

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: October 31, 2002, 13:25:48 : Search time 29.5 Seconds
(without alignments)
3864.331 Million cell updates/sec

Title: US-09-841-553-5
Perfect score: 3428
Sequence: 1 MGKALILVILVGLVGS.....YASTYGWADYQLKAVVYG 659

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_minc:*
- 8: sp_organelle:*
- 9: sp_plant:*
- 10: sp_virus:*
- 11: sp_vertebrate:*
- 12: sp_unclassified:*
- 13: sp_vertebrate:*
- 14: sp_virus:*
- 15: sp_rv:*
- 16: sp_bacteriophage:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	748.5	21.8	1245	Q9RL54	Q9RL54 streptomycin
2	688	20.1	1421	Q31788	Q31788 bacillus su
3	685.5	19.5	1102	P55684	P55684 streptomycin
4	667.5	19.5	1444	Q9KJ77	Q9KJ77 bacillus ha
5	664.5	19.4	1220	Q9LQAO	Q9LQAO streptomycin
6	658	19.2	1253	Q9FC06	Q9FC06 streptomycin
7	642.5	18.7	1412	Q9AER6	Q9AER6 thermomycin
8	615.5	18.0	1239	Q9FB24	Q9FB24 streptomycin
9	564	16.5	374	Q9F941	Q9F941 bacillus li
10	562	16.4	379	Q9FDF4	Q9FDF4 bacillus li
11	561	16.4	374	Q9F942	Q9F942 bacillus li
12	559	16.3	374	Q9F943	Q9F943 bacillus li
13	556	16.2	379	Q53521	Q53521 bacillus li
14	551	16.1	379	Q53300	Q53300 bacillus li
15	548	16.0	376	Q44684	Q44684 bacillus am
16	545	15.9	379	Q45301	Q45301 bacillus li

17	544	15.9	382	2	Q45522
18	543.5	15.9	310	2	Q9FDF3
19	543.5	15.9	1398	1	Q9P9L1
20	541	15.8	379	2	Q45299
21	540.5	15.8	310	2	Q9FDF2
22	536.5	15.7	310	2	Q9F7C2
23	534	15.6	382	2	Q87655
24	523	15.3	381	16	Q07613
25	522	15.2	378	2	Q45466
26	512	14.9	361	2	Q45521
27	506	14.8	621	2	Q53401
28	504	14.7	824	2	Q45464
29	503	14.7	621	2	Q9F486
30	501.5	14.6	379	2	Q66153
31	500.5	14.6	275	2	Q987J4
32	500.5	14.6	275	2	Q91L66
33	496.5	14.5	384	2	Q56365
34	496.5	14.5	757	16	Q9K656
35	496	14.5	374	2	Q45523
36	490.5	14.3	401	2	Q9AER8
37	490	14.3	379	2	Q45467
38	489	14.3	629	2	Q93JG8
39	488	14.2	799	16	Q9KEM1
40	481	14.0	422	1	Q977F5
41	460	13.4	715	2	P70765
42	457	13.3	397	2	P97097
43	456	13.3	640	2	Q93413
44	451.5	13.2	372	16	Q9KEL3
45	449	13.1	627	16	Q9RUD0

ALIGNMENTS

RESULT 1

Q9RL54 PRELIMINARY: PRT: 1245 AA.

AC Q9RL54: 01-MAY-2000 (T-EMBLrel. 13, Created)

DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)

DE PROBABLE SECRETED PEPTIDASE.

GN SCF51A.10

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

CC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.

CX NCBI_TaxID:1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA Murphy L., Harris D.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA Redenbach M., Kleser H.M., Denapante D., Eichner A., Cullum J., Kinashi H., Hopwood D.A.;

RL "A set of ordered cosmid and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RT Mol. Microbiol. 21:77-96(1996).

DR EMBL: AL121596; CAB56662.1;

DR RSP: P00782; ISUE.

DR MEROPS: S08.00A;

DR InterPro: IPR003137; PA.

DR InterPro: IPR000209; Peptidase_S8.

DR Pfam: PF02225; PA: 1.

DR Pfam: PF00082; Peptidase_S8; 1.


```
ID P95684 PRELIMINARY: PRT: 1102 AA.
AC P95684
DT 01-MAY-1997 (TEMBLrel. 03, Created)
DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE SUBTILISIN-LIKE PROTEASE.
OS Streptomyces albogriseolus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-1253.
RA MEDLINE=97144328; PubMed=9990295;
RX Suzuki M., Taguchi S., Yanada S., Kojima S., Miura K., Momose H.;
RT "A novel member of the subtilisin-like protease family from
RT Streptomyces albogriseolus."
RL J. Bacteriol. 179:430-438(1997).
DR EMBL: D83672; BAAL2040.1;
DR HSP: P00782; 2SPT.
DR MEROPS: S08.00A.1.
DR InterPro: IPR002860; BNR.
DR InterPro: IPR002029; Peptidase_S8.
DR Pfam: PF02012; BNR.2.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR Protease.
KW SEQUENCE.
SQ SEQUENCE 1102 AA; 114128 MW; F9E4AD2590F559E CRC64;

Query Match
Best Local Similarity 19.5%; Score 668.5; DB 2; Length 1102;
Matches 202; Conservative 71; Mismatches 228; Indels 127; Gaps 22;

Oy 109 GLTGCKALKSGVRIQEDYKVTYSALGLDESAAQVYNNVNLGYDGGSGITIGITDGT 168
Db 172 GTYDRTASGVARWLD-----GVRKASLDTSYGQITPKAWAGYDGGKGVKVLADTG 226

Oy 169 IDASHPLOGKVGWDFVNGRSPYDDHGHGTHVASIAAGTAASNGKYKGMAGAKLA 228
Db 227 VDATHPDLKGQVATASKNFTSAFT-TGDVVGHTHVASIAAGTGROSKYKGVAPGAKIL 285

Oy 229 GIKVLGADGSGSSTTIKGVENAVDNKDKYKINLSLGSQSSQSGTDSLSQAVNN-AW 287
Db 286 NCKVLDDAGFGDSDGLACHENAA-----AQGADIVNLSLGGMDTPE-TDPLEAAVDKLSA 340

Oy 288 DAGIVVCAAGNCPYTYTGSPAAASKVITVGAVDSDNIIASFSSRGP-TADGRKPEV 346
Db 341 EKGILFAIAGNEGPO--SIGSPGSADSLTVGAVDDKDLADESTSPRLGDGAYKPOL 398

Oy 347 VAPGVDIAPRASGTSNGTPIINDY---YTKASGTSNATPHVSGVLCALILQHPSTPDKY 403
Db 399 TAPGVDTIAASAKGNDIAKEVCEKPGACWTISGTSNATPHVAGAAALLKQHPENYAE 458

Oy 404 KTALE-ITADIAPKRIADIAGAGVNYKAIKY-----DD---Y 440
Db 459 KGALTASTKD-----GKVTFFQSGRQVQDKAITQVTAEPVSLSFQVQWPHADKPV 514

Oy 441 AKLTF-----TGSVAKGSATHT-----FDVSGATFVATLYNDTSSDIDLYVD 486
Db 515 KKUTYRNLTGEDVYTKLSTATGPKGAAPAGFTLGASTLTVA-----NGTASVDVTD 571

Oy 487 PNGNEVDYSYATYGFYKGVYNNPTAGTWTYKVSY-----KGAANYQ----- 529
Db 572 RLGAVDGTYSAVYATCAGOSVTAAREVEREVESVNTLVKLDORSKATANTMAYLSGL 631

Oy 530 -----VDVYVDSGLS-----QSGG-----GNPNPNPNPNPT 555
Db 632 TGLGKORSYAPTEADGAVKRVKGGYVLDA SVLYGADPETWRGADWLAQPKLDVTRNT 691

Oy 556 PTTDTQ-----TFTGSVNDYWDTSFTTNVNSGATKITGDLTDTSDTLDLKLDP 608
```

```
Db 692 VTVDARKAKPVKVTVPCKAKAQAQFADYTTIETNDNSAVSYGWLNSYGFSAHL----- 746
Oy 609 NGNLVDRSTSSSYEHVEYANPAGTWT 636
Db 747 -GPQITNGTLSSQW-NTHESNGAKAQYT 772

RESULT 4
O9KBJ7
ID O9KBJ7 PRELIMINARY: PRT: 444 AA.
AC O9KBJ7
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE INTRACELLULAR ALKALINE SERINE PROTEASE.
GN APRX OR BH1930.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11038132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji P., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001513; BAB05649.1;
DR HSP: O99405; 1MPT.
DR MEROPS: S08.00A.1.
DR InterPro: IPR001128; Cyt_P450.
DR InterPro: IPR002029; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR Protease; Complete proteome.
KW SEQUENCE.
SQ SEQUENCE 444 AA; 48916 MW; 3B05P81C53A21P2P CRC64;

Query Match
Best Local Similarity 19.5%; Score 667.5; DB 16; Length 444;
Matches 156; Conservative 80; Mismatches 122; Indels 67; Gaps 12;

Oy 35 VEKNYGLLTGCLFRKIQKLNPNNEISTVVFENHREKIAVRVLELMGAKVYVHIIPA 94
Db 57 VELNCGCFDQG-----SMTVEHLKHKQ-----COLHSEKIRC 92

Oy 95 IAADL---KVRDLVLSGLTGCKAKLSGVRIQEDYKVTYSALGL-LOESAQVMTYV 150
Db 93 VSANTPTSELQENLVCKDIR-----KIYLNREVHALLOTAVESAQAPEV 136

Oy 151 WNLG--YCGSITIGITDIDASHIPDLOGKVGWDFVNGRSPYDDHGHGTHVASIAA 208
Db 137 INGTETLTKGDTVIATVDTGCI-YPHIEDLEGRIKAFVDFVNGRSPYDDHGHGTHCAGDAA 195

Oy 209 GTGAASNGKYKCMAPKAGIKVLGADGSGSSTTIKGVENAVDNKDKY---GKIVNL 265
Db 196 NGASSDQYRGPAEPANVIGVYLNKOGMSLFSINQGVENCIGYNEEDPDPIHIIISM 255

Oy 266 SGSS---QSDGDTSLSOAVNNDAGIVYCVVAAGNSGPNYTYVCSPPAAASKVITV 322
Db 256 SLGGCALPYNEQEDPMVRIVEEAWNAGITVCVAAGNSGPDQGTIASPGVSEKIVT 315

Oy 323 D-----SNDTASFSSRGPTADGRKLPVWAPGVDTIAPRASCT-----SMCTPINDY 370
Db 316 DDDRTDREDDDYAPPSRGPTIYKPKPKDILAPGVNIVSLRSPNSFFYDKTKGSRVGS 375
```

```

QY 371 YTKASGSMATPHVSGVALILOAHPSWTPDKVKTALITETADIVAPKEIADIAYGAGRYN 430
DB 376 YTKASGSMATPCVAGVALILOAHPSWTPDKVKTALITETADIVAPKEIADIAYGAGRYN 433
QY 431 YKAI 435
DB 434 REGAI 438

RESULT 5
Q9LOAO PRELIMINARY; PRT; 1220 AA.
AC Q9LOAO;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE SECRETED PEPTIDASE.
GN SCC24.17C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Brown S.P., Harris D.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL EMBL; AL165003; CAB6111.1; -.
DR MEROPS; S08.00A; -.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR PRINTS; PR00723; Peptidase_S8; 1.
DR PROSITE; PS00136; SUBTILISIN.
DR PROSITE; PS00137; SUBTILISIN.
DR PROSITE; PS00138; SUBTILISIN.
DR PROSITE; PS00139; SUBTILISIN.
SQ SEQUENCE 1220 AA; 127311 MW; D9BD0018A7F31674 CRC64;

Query Match 19.4%; Score 664.5; DB 2; Length 1220;
Best Local Similarity 43.0%; Pred. NO. 1e-25;
Matches 161; Conservative 62; Mismatches 114; Indels 37; Gaps 14;

QY 118 SGVRFIOEDYKTVSAELEGIDESAAGVMAVYVNLGYDGSIGITIGITIGIDASHPOLQ 177
DB 184 AGVEGWLDGRYAD-----MAESNAQIGTPENAEAGTGRVTVAVLDGVDAGHPDLA 238
QY 178 GKVIGWVNVGRSPVDDRGHGHVASTAGTGAASNGKYKMGACAKLAGIKVLGADG 237
DB 239 GRAQSRSPICEEVA-DRHGHGHVTVSTVGGGAASDGRKGVAPONTLAVKGVLDDEG 297
QY 238 SGIISTIRKGVENAVDNDRKYKIVNLSIGSSQSSDGTSLGSAVNN-AMDAGIVCVVA 296
DB 298 FGSESEIAGMEWAARDVD---ADIVMSLSGSTPESDGTDPMAEAVNTLSRETGALFVIA 354
QY 297 AGNSGNTVYVTPGPAASKVITTVGAVDNDNTASFSSRGPP-ADGRKLPVAVAPGVDTIA 355
DB 355 AGNTGAPS-SIGSPGAADAALTGVGAVDSADQANWTSAGPRYGNALKPDLSPAGVGTILA 413

```

```

QY 356 PR---ASGTSMTIPINDYTKASGTSMTATPHVSGVALILOAHPSWTPDKVKTALITETAD 412
DB 414 ARSLEAESSG-----DYTSMDGTSMTATPHVAGVAALLAEHHPDWSGARLKDALMSTS- 465
QY 413 IVAPKEIADIAY--GAGRVVYKAIKYDDYAKLFTGSGVADKGSNTHTFDVSGATFVAT 470
DB 465 ----KELDVSAVQAGKRVSPAAVGAD-----VTATGS-ADLGYSPWPEADPE--VTRT 514
QY 471 LYWDGTGSSDIDL 484
DB 515 VAY-TNSTDDEL 527

RESULT 6
Q9FC06 PRELIMINARY; PRT; 1253 AA.
AC Q9FC06;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE SECRETED PEPTIDASE.
GN SC8A11.04C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL EMBL; AL391041; CAC01576.1; -.
DR HSP; Q99405; IMPT.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR PRINTS; PR00723; Peptidase_S8; 1.
DR PROSITE; PS00136; SUBTILISIN.
DR PROSITE; PS00137; SUBTILISIN.
DR PROSITE; PS00138; SUBTILISIN.
DR PROSITE; PS00139; SUBTILISIN.
SQ SEQUENCE 1253 AA; 130971 MW; AA69B417EEDB89 CRC64;

Query Match 19.2%; Score 558; DB 2; Length 1253;
Best Local Similarity 33.1%; Pred. NO. 2.3e-25;
Matches 195; Conservative 86; Mismatches 212; Indels 94; Gaps 24;

QY 13 VLGLVWGS-VAAAEKKVQVRNVKNYG-----LTPG----- 45
DB 49 VLTITGDRVTVGEGDAETVLSVTDPRHSGGARHVTGSDTYVYPDAVPYLGSGALD 108
QY 46 --LPRKIQKLNPN-----EISTVIVPENREKEIAVRVLELWGA-KYRVYHIIIPAIA 96
DB 109 ERLFNWTELLEDGVDARADELPLIVTDTAARSIGARTPE--GARRTRALLSIRGAAI 166
QY 97 ADLKVYDOLLVLSGLTG-----GKAKLSGVRFIOEDYKTVSAELEG-LDESAQV 145
DB 167 SAHESRAADPFWTSLTGTGDAAAGGSAKATSGRAGGAGTAKVLDGKVRATLSDTTAQI 226
QY 146 MATYVNLGYDGSIGITIGITIGIDASHPOLQKVICWDFVNGRSYPYDDHGHGHV 205

```

Db 227 GAPDVNSGGNTGGGVAVLDYDAGHPDFAGRIATAATASFVDDVDT-DRNGHGTHTAS 285
QY 206 TAAGTGAASNGKYKCHAPGAKLAGIKVLGADGSGSISTIKGVENAVDNKDKYKIVNL 265
Db 286 TVAGTGAASGCVKGVAPCASLHIGKVLNDSGSGODSWVLACHENAV--RDQHA-KIVSN 342
QY 266 SLGSSQSDGDTLSQAVNNH--DAGIVCVAAAGNSCPNTYTVGSPAASKVITVGA 322
Db 343 SLGDS-PTDGTDLSPAVN--WLSAETGALFYVAAGNSGPEAYTVGTTPAADAALTVGAV 399
QY 373 DSN-----DNLASFSRCP-TADGRKPEVYVAPGVIIAPRASGTSNGTPIINDYTKASGT 377
Db 400 NGPKGVQDVLADFSRGRVGDVNAVKPDLTAPGVGLAARSYAPEG---EGAYQSLSGT 456
QY 378 SNATPHVSGVGLILQAHPSWTPDKVKTALITETADIVAPKEIADIAGACRVVYKATKY 437
Db 457 SMATPHVAGAAALLAAEHPDWTGORKEALVGT--ACTORSPFDAGSGRVDVAAVR- 513
QY 438 DOYAKLTFTGSAVDKGSATHFTFVSGATFTATLYNDTGSSDI---DLYLYD---PNGNEV 492
Db 514 ---STLLASG---DAFAQHYPYTPGOTVRDVTYTNSGPAPVALDLALSPALPEG--- 564
QY 493 DYSTAYAYGFKVGYNTACTWTYKVSYKGAAN---YQVDVYSDGS 537
Db 565 -----LFTLSEAQVTPAHGTASGVYTHLDAEDNGAYATRLVASGA 607

RESULT 7

Q9AER6 ID Q9AER6 PRELIMINARY; PRT: 412 AA.
AC Q9AER6
DT 01-JUN-2001 (Tremblrel 17, Created)
DT 01-JUN-2001 (Tremblrel 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel 19, Last annotation update)
DE THERMICIN (SUBTILISIN-LIKE PROTEASE THERMICIN).
OS Thermoanaerobacter yonsei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Thermoanaerobacter.
OX NCBI_TaxID=111519;
RN [1]
RP SEQUENCE FROM N.A.
RA Jang H.J., Kim B.C., Pyun Y.R., Kim Y.S.;
RT "A novel subtilisin-like serine protease
RT yonseiensis KB-1: cloning, expression and biochemical properties.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Jang H.J., Kim B.C., Pyun Y.R., Kim Y.S.;
RT "Subtilisin-like protease, thermicin, from Thermoanaerobacter
RT yonsei.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY028704; AAK27733.1;
DR EMBL; AF305633; AAL09366.1;
DR HSSP; Q45670; 10BI.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; P00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Protease.
SQ SEQUENCE 412 AA: 44503 MW: E3CGAOFB1B1AID47 CRC64;

Query Match
Best Local Similarity 18.7%; Score 642.5; DB 2; Length 412;
Matches 159; Conservative 70; Mismatches 153; Indels 43; Gaps 13;

QY 27 KKVQVRNEKYNGLTGLGRFKIKLNPNEEISTVIVFENHREKEIAVRVLELGAQVR 86
Db 14 KEITLQKIDKR-----LLNKASYL--RSECVPAIYSLNLYETLKSII-EKGGGSIK 63
QY 87 YVWHIIPAIAADLKVROLLVSGTGKAKLSGVRFIOEDYKVTVAELEGDSAAQVM 146

Db 54 FEPIIKGNSVNLPCDKLKH-----ASIKGIHFAEDSLVKLQHLI-----ATQEI 110
QY 147 ATYVWN-LGYDSCGIGITIGIDGIDASHIPDL---OGKVIQWYDFVNGRSYYPYDDHGHGTH 202
Db 111 ASRKANDLGTYGKGTITLAPDTGI-YPHDFTKPKNRIIAFTDVVNGKKQPYDDNGHGH 169
QY 203 VASIAAGTGAASNGKYKCHAPGAKLAGIKVLGADGSGSISTIKGVENAVDNKDKYKIV 262
Db 170 VAGDAGMGYASNGKYKGVAPENIVAVKVLDSYGRGSSDILAGQWVLONKKNYIRI 329
QY 263 INLSGSSQS-SGDTLSQAVNNHAWDAGIVVVAAGNSCPNTYTVGSPAASKVITVGA 321
Db 230 VLSIGETPALPTFLDPLVRGVDTLWKNIGIVVVAAGNSCPNYSITSPGTSRAITVGA 289
QY 322 VD-----SDNIASFSSRGPTADGRKPEVYVAPGVVDIAPRASGTSNGTP----INDYY 371
Db 290 VDKRTDPDIEDDEVAKFSGRG--GPVLYKPDVYAPGVKIVSTASGNVPFGADEIMINKPY 347
QY 372 TRASGTSNATPHVSGVGLILQAHPSWTPDKVKTALITETADIVAPKEIADIAGACRVVY 431
Db 348 RSATGTSNATPHVAGAVALLLEKNSRLTNVEIKNLTKTATKI--NEAGLWTCGSGMINI 405
QY 432 YNAIK 436
Db 406 EEALK 410

RESULT 8

Q9FBZ4 ID Q9FBZ4 PRELIMINARY; PRT: 1239 AA.
AC Q9FBZ4
DT 01-MAR-2001 (Tremblrel 16, Created)
DT 01-MAR-2001 (Tremblrel 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel 19, Last annotation update)
DE PUTATIVE SECRETED PEPTIDASE.
GN SC8A11.16C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteri; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdano A.M., Parkhill J., Bartell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000031; PubMed=843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinsahl H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL391041; CAC01588.1;
DR HSSP; Q99405; 1MPT.
DR MEROPS; S08.00A;
DR InterPro; IPR001137; PA.
DR InterPro; IPR00209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; P00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
SQ SEQUENCE 1239 AA: 128505 MW: 8F5E9AC68EB1260A CRC64;

Query Match
Best Local Similarity 18.0%; Score 615.5; DB 2; Length 1239;
Pred. No. 3.2e-23;

Matches 182; Conservative 72; Mismatches 196; Indels 119; Gaps 19;

QY 100 KYRDLVLSG-----LIGGKAKLSGVRFIOED-----YKVTVSAALE 136
 Db 161 KYRDLASVQGAALSADPARAAEFWSYVTEGPGAGGVDSGTDAAFRGGVAHIWLDAPVE 220
 QY 137 G-LDSEAQVATVWNLGVDGSGITIGLIDTGDASHDPLOGKVIHGVDFVNGRSYPYD 195
 Db 221 ADLADSTAQIGAPRANAGNTGGVEVAVLDTGVDAGHPDLADRRAQSFVDFENID-D 279
 QY 196 DGHGTHVASTAAGTGAASNGRYKMGAPCAKAGIKVLGADSGSISITIKGVEMAVDNK 255
 Db 280 ROGHCTHVNSTIAGTGAASAGKEKGVAPCARLSIGKVLNDSGRGQISMTLAAMEWAAVER 339
 QY 256 DKYGIKVINLSIGSSQSDGTSLSQAVNN-AMDAGTVWCVAAGNSGPNVTYVGSAAAS 314
 Db 340 H---AKIVNMSLGSGEUSGSDPMSRAVDRLSAQTGALFVVRAGNGG-EAGSIGAPCVAI 395
 QY 315 KYTVGAVDSNDNTASFSRGPADGRLKPEVAVPVDIITAPRASGTSNGTPIINDYTKA 374
 Db 396 SALTGVADATDTLAPTSSQGPVYDGLAKPEITAPGVGLLAANSFPAAG---NGATQSL 452
 QY 375 SGTSMATPHVSGVALILQARHPSWTPDKYKTALITADIVAPKEIADIAY-----GAGRV 429
 Db 453 SGTSMATPHVAGAAALLAARPDLSGALK-----DVLASSSHRTPRYDAFQAGSGRV 505
 QY 430 NVYKATKYDDVAKLTFETGVSADKGSATHTFDVSGATFVTATLYWDTGSSDIDLILYDPNG 489
 Db 506 DVDAVRAQVYASATAPGSSGPPVRLVTVTNTGAAVTL-----547
 QY 490 NEVDYSYATYGFYKGVYNTACTVTVKVSYGKAANYQVGVDSLSGSSGNGNPN 549
 Db 548 ---ELSAATHAPE--GVFLASRVTPA---HGTAD--VTLIDGSSAG-----590
 QY 550 PNPPTPTDTDTFTGVSNDYMDTSDFTNNVNSGATKINGDLTFTDSYNDLKLILYDPN 609
 Db 591 -----RAYSGQIL-----ATDADARNVAHTAVS-AGPVR-----HKLTVHFKDAD 629
 QY 610 GNLVDRSTSSNSYEHVEVANPOTWFL 638
 Db 630 G-----NVPQGVFDLL 640

RESULT 9
 Q9F941 PRELIMINARY; PRT; 374 AA.
 AC Q9F941;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE SUBTILISIN PRECURSOR (EC 3.4.21.62) (FRAGMENT).
 GN SUBC.
 OS Bacillus licheniformis.
 OC Bacteria, Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1402;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC12759;
 RX MEDLINE=20561724; PubMed=11109488;
 RA Evans K.L., Crowder J.S., Miller E.S.;
 RT "Subtilisins of Bacillus spp. hydrolyze keratin and allow growth on
 feathers".
 RL Can. J. Microbiol. 46:1004-1011(2000).
 DR EMBL; AF205191; AG31028.1; .
 DR HSP; P00780; ICSE.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 KW Signal; Hydrolase.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 >374 SUBTILISIN.

FT NON_TER 374 374
 SQ SEQUENCE 374 AA; 38423 MW; 640142EB813D0E50 CRC64;

Query Match 16.5%; Score 564; DB 2; Length 374;
 Best Local Similarity 35.3%; Pred. No. 2.8e-21;
 Matches 153; Conservative 62; Mismatches 145; Indels 74; Gaps 15;

QY 3 GLKALLILVILGLVGVSAVAAPKKEKVEQVRNVEKNYGLLTGCLFRKTKOKLNPNEEISTV 62
 Db 10 GMLTAFMLVETMAESDSASAAQPAK-----NVERNY-----I 41
 QY 63 IVTENH-REKETAVRVELEAKCAKRVYVYHITPAIADLKVRDLDLVISGLTGCKAKLSGR 121
 Db 42 VGRFSGVATASVYKDIIRKESGKVKQDFRIINRAKAKLDKEALKEVK-----NDPVA 94
 QY 122 FIOEDYKVTVSABLEGDESAQAQVMAHYVWNLGTDGSGITIGLIDTGDASHDPLOGKVI 181
 Db 95 YVEEDH--VAHALAQTVPGYGIPLKADKVRQGFNGKAVRYAVLDTGIOASHPDL--NVV 150
 QY 182 GWDFYNGRSYPYDDHGHGTHVASIAAGTGAASNGK--YKMAPGAKLAGIKVLGADSG 239
 Db 151 GGASFVAGEAYNTDCNGHGHV---AGTVAALDNTTCVLGVAPSVSLXAVKVLNSSGSG 206
 QY 240 SISTIIRKGVEMAVDNKDKYIKVINLSIGSSQSDGTSLSQAVNNAMDAGIVVCVAAGN 299
 Db 207 SYSGIVSGIENATN---GMDVINKSLG---GASGSTAMKQAVDNAYARGVWVVAAGN 259
 QY 300 SGP--NTYTVGSPAASKVTITVCAVDSNDNIASFSSRGPTADGRLKPEVAVGVDIAPR 357
 Db 260 SGSSGNTNTIGYPAKYDVIAGVAVDSNRSRFSVSG-----AELEVNAPG-----306
 QY 358 ASGTSMTGTPINDYTKASGTSMAHPVSGVALILQARHPSWTPDKYKTALITADIVAPK 417
 Db 307 -AGVSYTPNTYAT-LNGTSMASPHVAGAAAILLSRHPNLSASQVRNLSSTATYLG--362
 QY 418 EIADIYAGRVNV 431
 Db 363 --SFYIGKGLINV 374

RESULT 10
 Q9FDF4 PRELIMINARY; PRT; 379 AA.
 AC Q9FDF4;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE KERA.
 OS Bacillus licheniformis.
 OC Bacteria, Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1402;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OWU 1411T;
 RA Goldstein G., Wagner E.K., Lee Y.E., Masisak E.S., Oh J.;
 RT "Nucleotide sequence of kera gene encoding a keratinolytic protease of
 Bacillus licheniformis OWU 1411T".
 RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF282893; AAG00492.1; .
 DR HSP; P00780; ICSE.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 SQ SEQUENCE 379 AA; 38904 MW; DC65F6A93EBC69CC CRC64;

Query Match 16.4%; Score 562; DB 2; Length 379;
 Best Local Similarity 34.5%; Pred. No. 3.6e-21;
 Matches 151; Conservative 64; Mismatches 148; Indels 74; Gaps 14;

QY 3 GLKALLILVILGLVGVSAVAAPKKEKVEQVRNVEKNYGLLTGCLFRKTKOKLNPNEEISTV 62
 Db 10 GMLTAFMLVETMAESDSASAAQPAK-----NVERNY-----I 41

Db 10 GMLTALMLVTFMAFSASAAQPAK-----NVEKDY-----I 41
 QY 63 IVEENH-REKETAVRLELAKGAKRYVYHIIPAIADLVKRDLLVISGLTGKAKLSGVR 121
 Db 42 VGFSGVKTSVSKKDIRESGKVDKQFRIINAARAKLDKEALKEVK-----NDPDA 94
 QY 122 FIOEDYKVTYSAELEGDESAAQVMATYVWNLGYDGGSTIGITDGTIDASHPDLOGKVI 181
 Db 95 YVEEDH--VAHALAOTVPYIGPLIKADKVOAQGFKGANKVAVLDGTQASHPDL--NVV 150
 QY 182 GWDVDFVNGRSYDDHGHGTHVASTAGTGAASNGK--YKGMAPCAKLAGIKVLGADGSG 239
 Db 151 GGASPFVAGEAYNTDNGHGHV----AGTVAALDNTTGVLGVPVAVSLYAVKVLNSSGSG 206
 QY 240 SISTIITKGVENAVDNKDKYGIKVINLSGLSSQSSDGTSLQAVNNAMDAGIYVYVAAGN 299
 Db 207 SYSGIVSGIEMATTN----GMDVINSLG---CPGSGTAMKQAVDNAYARGVVAAGN 259
 QY 300 SGP--NTYTVGSPAAASKVITVGVADSDNINIASFSSRGPTADGRKPEVYVAPGVDIAPR 357
 Db 260 SGSSGNTNTIGYPAKYDSVIAVGADSDNSNRASFSVG-----AELEYMAPGAGYSTY 313
 QY 358 ASGTSNGTPTINDYTKASGTSKATPHVSGVGLIQAHPSTWPKVKTALITETADIVAPK 417
 Db 314 PTST-----YATLNGTSMASPHVAGAAALILSKHPNLSASQVRNRLSSTATYLG-- 362
 QY 418 EIADIAYAGRVNYKA 434
 Db 363 --SSFYVYKGLINVEAA 377

RESULT 11

Q9F942 ID Q9F942 PRELIMINARY: PRT: 374 AA.
 AC Q9F942
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE SUBTILISIN PRECURSOR (EC 3.4.21.62) (FRAGMENT).
 GN SUBC.
 OS Bacillus licheniformis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1402;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIMB 10689;
 RX MEDLINE=20561724; PubMed=11109488;
 RA Evans K.L., Crowder J.S., Miller E.S.;
 RT "Subtilisins of Bacillus spp. hydrolyze keratin and allow growth on
 featherers.";
 RL Can. J. Microbiol. 46:1004-1011(2000).
 DR EMBL; AF205190; AAG31027.1;
 DR HSSP; P00780; ICSE.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 KW Signal; Hydrolase.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 >374 SUBTILISIN.
 FT NON_TER 374 374
 SQ SEQUENCE 374 AA: 38390 MW: 86143214B42C5262 CRC64;

Query Match 16.4%; Score 561; DB 2; Length 374;
 Best Local Similarity 35.0%; Pred. No. 4e-21;
 Matches 152; Conservative 64; Mismatches 144; Indels 74; Gaps 15;

QY 3 GLKALILVILVGLVGVGVAAPKVEQVRNVEKNYGLLTGTLGFRKIQKLNNEISYV 62
 Db 10 GMLTALMLVTFMAFSASAAQPAK-----NVEKDY-----I 41
 QY 63 IVEENH-REKETAVRLELAKGAKRYVYHIIPAIADLVKRDLLVISGLTGKAKLSGVR 121

Db 42 VGFSGVKTSVSKKDIRESGKVDKQFRIINAARAKLDKEALKEVK-----NDPDA 94
 QY 122 FIOEDYKVTYSAELEGDESAAQVMATYVWNLGYDGGSTIGITDGTIDASHPDLOGKVI 181
 Db 95 YVEEDH--VAHALAOTVPYIGPLIKADKVOAQGFKGANKVAVLDGTQASHPDL--NVV 150
 QY 182 GWDVDFVNGRSYDDHGHGTHVASTAGTGAASNGK--YKGMAPCAKLAGIKVLGADGSG 239
 Db 151 GGASPFVAGEAYNTDNGHGHV----AGTVAALDNTTGVLGVPVAVSLYAVKVLNSSGSG 206
 QY 240 SISTIITKGVENAVDNKDKYGIKVINLSGLSSQSSDGTSLQAVNNAMDAGIYVYVAAGN 299
 Db 207 SYSGIVSGIEMATTN----GMDVINSLG---GASGSTAMKQAVDNAYARGVVAAGN 259
 QY 300 SGP--NTYTVGSPAAASKVITVGVADSDNINIASFSSRGPTADGRKPEVYVAPGVDIAPR 357
 Db 260 SGSSGNTNTIGYPAKYDSVIAVGADSDNSNRASFSVG-----AELEYMAPGAGYSTY 313
 QY 358 ASGTSNGTPTINDYTKASGTSKATPHVSGVGLIQAHPSTWPKVKTALITETADIVAPK 417
 Db 307 -AGVYSTPTWTYAT-LNGTSMASPHVAGAAALILSKHPNLSASQVRNRLSSTATYLG-- 362
 QY 418 EIADIAYAGRVNY 431
 Db 363 --SSFYVYKGLIN 374

RESULT 12

Q9F943 ID Q9F943 PRELIMINARY: PRT: 374 AA.
 AC Q9F943
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE SUBTILISIN PRECURSOR (EC 3.4.21.62) (FRAGMENT).
 GN SUBC.
 OS Bacillus licheniformis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1402;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIMB 6816;
 RX MEDLINE=20561724; PubMed=11109488;
 RA Evans K.L., Crowder J.S., Miller E.S.;
 RT "Subtilisins of Bacillus spp. hydrolyze keratin and allow growth on
 featherers.";
 RL Can. J. Microbiol. 46:1004-1011(2000).
 DR EMBL; AF205189; AAG31026.1;
 DR HSSP; P00780; ICSE.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 KW Signal; Hydrolase.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 >374 SUBTILISIN.
 FT NON_TER 374 374
 SQ SEQUENCE 374 AA: 38396 MW: 6048C17B985413B1 CRC64;

Query Match 16.3%; Score 559; DB 2; Length 374;
 Best Local Similarity 35.0%; Pred. No. 5e-21;
 Matches 152; Conservative 63; Mismatches 145; Indels 74; Gaps 15;

QY 3 GLKALILVILVGLVGVGVAAPKVEQVRNVEKNYGLLTGTLGFRKIQKLNNEISYV 62
 Db 10 GMLTALMLVTFMAFSASAAQPAK-----NVEKDY-----I 41
 QY 63 IVEENH-REKETAVRLELAKGAKRYVYHIIPAIADLVKRDLLVISGLTGKAKLSGVR 121
 Db 42 VGFSGVKTSVSKKDIRESGKVDKQFRIINAARAKLDKEALKEVK-----NDPDA 94

QY 229 GIKVLGADGSGSTIIKGVENAVDNKOKYIKVYNLSLGGSSQSDGTDLSQAVNAMD 288
 Db 196 AVKVLNLSGSGSTSAIVSGIENATTT-----GMDYNNLSLGGASVS---TAMKQAVDHAYA 248
 QY 289 AGIYVVAAGNSGP--NTYVGSPPAAASKVITYGAVDSNDNTASFSSRGPTADGRKPEV 346
 Db 249 RGAVVVSAGNSGSSGNTIGYPKYDSVIAGCAVDSNRSASFSSVG-----AELRV 302
 QY 347 VAPGVDIAPRAGTSMGTPIINDYTKASGTSMATPHVSGVAGLILQAHPSWTPDKVKTA 406
 Db 303 WAPG-----AGVSYTPTNYAT--LNGTSMASPHVAGAAALILSKHPNLSASQVRTR 353
 QY 407 LIETADIVAPKETADIAACAGRVNYKA 434
 Db 354 LSRTATYLG-----SFSYGRGLINVEAA 377

RESULT 15
 Q44684

ID Q44684 PRELIMINARY; PRT; 376 AA.
 AC Q44684;
 DT 01-NOV-1996 (TREMELrel. 01, Created)
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
 DE SUBTILISIN PRECURSOR.
 OS Bacillus amyloliquefaciens.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_taxid=1390;
 RP [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=84069812; PubMed=6316278;
 RA Wells J.A., Ferrari E., Henner D.J., Estell D.A., Chen E.Y.;
 RT "Cloning, sequencing, and secretion of Bacillus amyloliquefaciens
 subtilisin in Bacillus subtilis."
 RL Nucleic Acids Res 11:7911-7925(1983).
 DR EMBL: X00165; CAA24990.1;
 DR HSSP: P00782; 1SUP.
 DR MEROPS: S08.034;
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KW signal.
 FT SIGNAL
 FT CHAIN 1 27 POTENTIAL.
 FT CHAIN 103 376 POTENTIAL.
 SQ SEQUENCE 376 AA; 39426 MW; BA6FD7A11BDBBEE CRC64;

Query Match 16.0%; Score 548; DB 2; Length 376;
 Best Local Similarity 34.8%; Pred. No. 1.8e-20;
 Matches 154; Conservative 65; Mismatches 142; Indels 82; Gaps 17;
 QY 4 LKALILVILGLVGVSAAPKKEVQVNVKYNGLTLPGLFRKIOKLNPNEEISTVI 63
 Db 2 ISLLFALALIFTNAGSTSSA---QAAGKSNKEKKY-----IV 36
 QY 64 VFENHREKELAVR---VLELCAKRVYVYHIIPAIADLK---VRDLLVIGSLTGGKAKL 117
 Db 37 GFKQTMTGTHSAHKVKVISEKGGVKQFKYVDAASATLNKAVKEL-----KKD 86
 QY 118 SCVRFIOEDYKVTVALEGLDESAQAQVATYVYNNLGVDSGTTIGITDGLDASHPDQ 177
 Db 87 PSVAYVEEDH--VAHAYAQSPVYVQIKAPALHSGYTGSNVKVAVIDSCIDSHFDL- 143
 QY 178 KVIYGVDFVNGRSPY--DDHGHGTHVASTAGTGAASNGK--YKGMAPGAKLAGIKVLG 234
 Db 144 -KVAGASVPSVETPNFPODNNSHGTHV-----ACTVAALNNSIGYLVGPSASLIYAVKVLG 198
 QY 235 ADGSGSSTIIKGVENAVDNKOKYIKVYNLSLGGSSQSDGTDLSQAVNAMDAGIYVC 294
 Db 199 ADGSGQSWIINGIEIAIAN-----NMDVINNSLG---GPSGSAALKAAYDKAVASGVYV 251

QY 295 VAAGNSGP--NTYVGSPPAAASKVITYGAVDSNDNTASFSSRGPTADGRKPEVAPGVD 352
 Db 252 AAAGREGTSGSSSTYGVYGPYPSVIAVAGAVDSNORASFSSVYCPELD-----VWAPGV 305
 QY 353 IIAPRASGTSMTGTPINDYTKASGTSMATPHVSGVAGLILQAHPSWTPDKVKTALETAD 412
 Db 306 I-----QSTLPCNKYGYNGTSMASPHVAGAAALILSKHPNNTQVRSLENTT- 355
 QY 413 IVAPKEIAD-IAYGAGRVNYKA 434
 Db 356 ----TKGSEFYTGKGLINVOAA 374

Search completed: October 31, 2002, 13:31:18
 Job time : 33.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:25:28 ; Search time 10 Seconds
(without alignments)
2551.616 Million cell updates/sec

Title: US-09-841-553-5

Perfect score: 3428
Sequence: 1 MKGLKALILVILVGLVGS.....YAYSTGYADYOLKAVVYG 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	555	16.2	382	1 SUBT_BACAM	P00782 bacillus am
2	554	16.2	379	1 SUBT_BACLI	P00780 bacillus li
3	540.5	15.8	1398	1 PLS_PVRPU	P2186 pyrococcus
4	537.5	15.7	380	1 ELYA_BACAO	P41362 bacillus al
5	537.5	15.7	380	1 ELYA_BACCS	P00781 bacillus li
6	535	15.6	274	1 SUBD_BACLI	P35835 bacillus li
7	524	15.3	361	1 SUBN_BACNA	P04189 bacillus su
8	523	15.2	381	1 SUBT_BACSA	P00783 bacillus su
9	520	15.2	381	1 SUBT_BACSA	P29142 bacillus st
10	520	15.2	381	1 SUBT_BACSA	P29141 bacillus ha
11	508	14.8	361	1 ELYA_BACHD	P29141 bacillus ha
12	505.5	14.7	806	1 SUBV_BACSU	P07318 bacillus pu
13	500.5	14.6	275	1 SUBT_BACPU	Q95405 bacillus sp
14	494.5	14.4	269	1 PRTM_BACSP	P29599 bacillus le
15	493.5	14.4	269	1 SUBB_BACLE	P29600 bacillus le
16	493.5	14.4	269	1 SUBS_BACLE	Q45670 bacillus sp
17	491.5	14.3	401	1 THES_THEVU	P04072 thermoactin
18	486.5	14.3	279	1 THES_BACSP	P58502 pyrococcus
19	486.5	14.2	378	1 ELYA_BACSP	P16397 bacillus su
20	481	14.0	422	1 TKSU_PYRKO	P08394 thermus aqu
21	452.5	13.2	1433	1 SUBF_BACSU	P16588 vibrio algi
22	428.5	12.5	513	1 AQLI_THEAO	P16396 bacillus su
23	426.5	12.4	534	1 PROA_VTBAL	P80146 thermus sp
24	416	12.4	645	1 SUBE_BACSU	P29143 halophilic
25	409	11.9	408	1 SEPR_THESR	P29143 halophilic
26	401	11.7	530	1 HLY_HALI7	P29143 halophilic
27	400	11.7	319	1 ISPI_BACSU	P29143 halophilic
28	390	11.4	420	1 SUBT_BACSU	P29143 halophilic
29	386.5	11.3	467	1 ISPE_SCHPO	P40903 schizosacch
30	380	11.1	595	1 BPRX_BACNO	P42780 bacteroides
31	379.5	11.1	635	1 PRTB_YEAST	P09232 saccharomyc
32	379	11.1	894	1 WPRB_BACSU	P54423 bacillus su
33	372.5	10.9	326	1 ISP_PAPEO	P29139 paenibacill

RESULT 1
SUBT_BACAM
ID AC P00782; STANDARD; PRT; 382 AA.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Subtilisin BPN' precursor (EC 3.4.21.62) (Subtilisin Novo) (Alkaline
DE protease).
GN APR.
OS Bacillus amyloliquefaciens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23844;
RX MEDLINE=8506739; PubMed=6090391;
RA Vasantha N., Thompson L.D., Rhodes C., Banner C., Nagle J.,
RA Filipula D.;
RT Genes for alkaline protease and neutral protease from Bacillus
RT amyloliquefaciens contain a large open reading frame between the
RT regions coding for signal sequence and mature protein.;
RL J. Bacteriol. 159:811-819(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84069812; PubMed=6316278;
RA Wells J.A., Ferrari E., Henner D.J., Estell D.A., Chen E.Y.;
RT "Cloning, sequencing, and secretion of Bacillus amyloliquefaciens
RT subtilisin in Bacillus subtilis.;
RL Nucleic Acids Res. 11:7911-7925(1983).
RN [3]
RP SEQUENCE OF 108-382.
RX MEDLINE=68086682; PubMed=6065094;
RA Markland F.S., Smith E.L.;
RT "Subtilisin BPN. VII. Isolation of cyanogen bromide peptides and the
RT complete amino acid sequence.;
RL J. Biol. Chem. 242:5158-5211(1967).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=72035041; PubMed=4399039;
RA Alden R.A., Wright C.S., Kraut J.;
RT "A hydrogen-bond network at the active site of subtilisin BPN'";
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 257:119-124(1970).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF COMPLEX WITH INHIBITOR.
RX MEDLINE=85033707; PubMed=6387152;
RA Hirono S., Akagawa H., Mitsui Y., Iitaka Y.;
RT "Crystal structure at 2.6-A resolution of the complex of subtilisin
RT BPN' with streptomyces subtilisin inhibitor.;
RL J. Mol. Biol. 178:389-413(1984).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF MUTANT.
RX MEDLINE=90057412; PubMed=2684274;
RA Pentoliano M.W., Whitlow M., Wood J.F., Dodd S.W., Hardman K.D.,
RA Rollence M.L., Bryan P.N.;

ALIGNMENTS

34 372 10.9 533 1 PEPC_ASPNG
35 371.5 10.8 580 1 PROA_XANCP
36 370.5 10.8 388 1 CUDP_METAN
37 369 10.8 387 1 PRTR_TRIAL
38 365 10.6 478 1 YSP3_YEAST
39 364.5 10.6 321 1 ISP_BACCS
40 361 10.5 384 1 PRTK_TRIAL
41 360.5 10.5 536 1 SPML_MAGGR
42 347.5 10.1 402 1 ALP_CEPAC
43 343 10.0 293 1 PRTE_TRIAL
44 340.5 9.9 491 1 YCT5_YEAST
45 332.5 9.7 1902 1 P2P_LACPA

P33295 aspergillus
P23314 xanthomones
P29138 metarhizium
P23653 tritirachiu
P25036 saccharomyc
P29140 bacillus cl
P06873 tritirachiu
P58371 magnaporthe
P29118 cephalospor
P20015 tritirachiu
P25381 saccharomyc
Q02470 lactobacill

```
DR DR PIR: A00570; UCSB.
```

```
DR DR PDB: 1S01; 15-OCT-90.
```

```
DR DR PDB: 1S02; 15-JAN-92.
```

```
DR DR PDB: 1SBH; 07-DEC-95.
```

```
DR DR PDB: 1SBH; 07-DEC-95.
```

```
DR DR PDB: 1SBH; 31-JAN-94.
```

```
DR DR PDB: 1SBH; 31-MAY-84.
```

```
DR DR PDB: 1SBH; 31-MAY-84.
```

```
DR DR PDB: 1SIB; 31-OCT-93.
```

```
DR DR PDB: 1SIC; 15-APR-93.
```

```
DR DR PDB: 1SIC; 31-JAN-94.
```

```
DR DR PDB: 1SIC; 31-JAN-94.
```

```
DR DR PDB: 1SNT; 15-JAN-93.
```

```
DR DR PDB: 1SPB; 15-OCT-95.
```

```
DR DR PDB: 1STI; 15-JUL-91.
```

```
DR DR PDB: 1STJ; 15-JUL-91.
```

```
DR DR PDB: 1SUA; 14-JAN-98.
```

```
DR DR PDB: 1SUB; 31-JAN-94.
```

```
DR DR PDB: 1SUC; 31-JAN-94.
```

```
DR DR PDB: 1SUD; 31-JAN-94.
```

```
DR DR PDB: 1SUE; 14-OCT-98.
```

```
DR DR PDB: 1SUP; 14-NOV-95.
```

```
DR DR PDB: 1XG9; 12-NOV-97.
```

```
DR DR PDB: 1XQ9; 31-DEC-97.
```

```
DR DR PDB: 1AQN; 14-JAN-98.
```

```
DR DR PDB: 1IAA; 11-JUL-96.
```

```
DR DR PDB: 1YBJ; 11-JUL-96.
```

```
DR DR PDB: 1YJC; 11-JUL-96.
```

```
DR DR PDB: 1AZQ; 23-APR-98.
```

```
DR DR MEROPS; S08.001;
```

```
DR DR InterPro; IPRA00209; Peptidase_S8.
```

```
DR DR Pfam; PF00008; Peptidase_S9; 1.
```

```
DR DR PRINTS; PR00723; SUBTILISIN.
```

```
DR DR PROSITE; PS00136; SUBTILASE ASP; 1.
```

Db 144 SSSPDL--KVAGSAYPSETNPFODNNSGTHV---AGTVAALNNSIGVLGVPASAL 197
 QY 228 AGIKVLGADSGSISIIIGVWVADNDKYGKIVNLISGSSQSDGTDLSOAVNNAM 287
 Db 198 YAVKVLGADSGGOVSHIINGIEMATN---NMDVINKSLG---QPSGSAALKAAYDKAV 250
 QY 288 DAGIVVCVAGNSGP--NNTYVGSPPAAASKVITVAVGAVDSNDNIASFSSRGPTADGRKPE 345
 Db 251 ASGVVYVAAAGNGTSGSSSTVPGKYPSPVIAVAVGAVDSNQRASFSSVGPFLD----- 304
 QY 346 VVAVGVIIAPRASGTMGTINDYTKASGTSMTATPHVSGVGLIQLQAPSPMTPKVKT 405
 Db 305 VNAFGVSI-----OSTLPGNKYGAYNGTSMASPHVAGAAALILSKHPNTNTQVRS 355
 QY 406 ALIETADIVAPKEIAD-IAYGAGRVNYKA 434
 Db 356 SLENTT-----TLGDSFYVKGGLINVOAA 380
 RESULT 2
 ID SUBT_BACLI STANDARD: PRT; 379 AA.
 AC P00780:
 DT 21-JUL-1986 (Rel. 01. Created)
 DT 01-AUG-1988 (Rel. 08. Last sequence update)
 DT 16-OCT-2001 (Rel. 40. Last annotation update)
 DE Subtilisin Carlsberg precursor (EC 3.4.21.62).
 GN APR.
 OS Bacillus licheniformis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1402;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIB 6816;
 RX MEDLINE=86093688; PubMed=3001653;
 RA Jacobs M., Eliasson M., Uhlen M., Flock J.-I.;
 RT "Cloning, sequencing and expression of subtilisin Carlsberg from
 RT Bacillus licheniformis";
 RL Nucleic Acids Res. 13:8913-8925(1985).
 RN [2]
 RP SEQUENCE OF 105-379
 RX MEDLINE=68234702; PubMed=4967581;
 RA Smith E.L., Delange R.J., Evans W.H., Landon M., Markland F.S.;
 RT "Subtilisin Carlsberg. V. The complete sequence; comparison with
 RT subtilisin BPN'; evolutionary relationships";
 RL J. Biol. Chem. 243:2184-2191(1968).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF VARIANT WITH SELENOCYST-325.
 RX MEDLINE=93291170; PubMed=8512925;
 RA Syed R., Wu Z.P., Hogle J.M., Hilvert D.;
 RT "Crystal structure of selenosubtilisin at 2.0-A resolution";
 RL Biochemistry 32:6157-6164(1993).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 105-379.
 RX MEDLINE=98087517; PubMed=9425066;
 RA Scoll V.S., Eger B.T., Hynes R.C., Martichonok V., Jones J.B.,
 RA Pai E.F.;
 RT "Differences in binding modes of enantiomers of 1-acetamido boronic
 RT acid based protease inhibitors: crystal structures of gamma-
 RT chymotrypsin and subtilisin Carlsberg complexes";
 RL Biochemistry 37:451-462(1998).
 CC -1- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
 CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
 CC for peptide bonds, and a preference for a large uncharged residue
 CC in P1. Hydrolyses peptide amides.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- BIOTECHNOLOGY: Used as a detergent protease. Sold under the name
 CC Alcalase by Novozymes.
 CC -1- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
 CC SPOREFORMATION, AND MANY MUTATIONS WHICH BLOCK SPOREFORMATION AT EARLY

CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
 CC IS NOT NECESSARY FOR NORMAL SPOREFORMATION.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
 CC or send an email to license@ebi.ac.uk).
 CC EMBL; X03341; CAB56500.1;
 DR PIR; A00969; SUBSC.
 DR PIR; A24111; SUBSCL.
 DR PDB; 1CSE; 15-OCT-89.
 DR PDB; 1SCA; 31-JAN-94.
 DR PDB; 1SCB; 31-JAN-94.
 DR PDB; 1SCD; 31-JAN-94.
 DR PDB; 1SBC; 15-JAN-95.
 DR PDB; 2SEC; 15-JAN-95.
 DR PDB; 1SCN; 31-AUG-94.
 DR PDB; 1AE4; 16-JUN-97.
 DR PDB; 1SEL; 31-OCT-93.
 DR PDB; 1VSB; 18-MAR-98.
 DR PDB; 3VSB; 25-MAR-98.
 DR PDB; 1AVT; 25-MAR-98.
 DR PDB; 1AV7; 01-APR-98.
 DR PDB; 1BE6; 14-OCT-98.
 DR PDB; 1BE8; 13-JAN-99.
 DR PDB; 1BPK; 18-NOV-98.
 DR PDB; 1BFU; 18-NOV-98.
 DR MEROPS; S08.001;
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PR00721; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Sporulation; Serine protease; Zymogen; Signal;
 KW 3D-structure.
 FT SIGNAL 1 29 POTENTIAL.
 FT PROPEP 30 105
 FT CHAIN 105 379 SUBTILISIN CARLSBERG.
 FT ACT_SITE 137 137 BY SIMILARITY.
 FT ACT_SITE 168 168 BY SIMILARITY.
 FT ACT_SITE 325 325 BY SIMILARITY.
 FT CONFLICT 207 207 T -> S (IN REF. 2).
 FT CONFLICT 233 233 P -> A (IN REF. 2).
 FT CONFLICT 262 265 SSGN -> NSGS (IN REF. 2).
 FT CONFLICT 316 316 S -> N (IN REF. 2).
 FT TURN 111 111
 FT HELIX 112 115
 FT TURN 116 117
 FT HELIX 118 122
 FT TURN 123 125
 FT TURN 129 130
 FT STRAND 132 137
 FT TURN 142 143
 FT TURN 145 146
 FT STRAND 149 154
 FT TURN 157 158
 FT TURN 161 162
 FT HELIX 168 177
 FT TURN 180 191
 FT STRAND 193 198
 FT TURN 202 203
 FT HELIX 208 220
 FT TURN 221 222
 FT STRAND 225 228
 FT STRAND 232 232
 FT HELIX 237 248

Query Match 15.9%; Score 540.5; DB 1; Length 1398;
 Best Local Similarity 22.5%; Pred. No. 4.6e-23;
 Matches 238; Conservative 103; Mismatches 251; Indels 465; Gaps 38;

FT CARBOHYD 929 929 N-LINKED (POTENTIAL).
 FT CARBOHYD 1048 1048 N-LINKED (POTENTIAL).
 FT CARBOHYD 1056 1056 N-LINKED (POTENTIAL).
 FT CARBOHYD 1084 1084 N-LINKED (POTENTIAL).
 FT CARBOHYD 1117 1117 N-LINKED (POTENTIAL).
 FT CARBOHYD 1133 1133 N-LINKED (POTENTIAL).
 FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
 FT CARBOHYD 1148 1148 N-LINKED (POTENTIAL).
 FT CARBOHYD 1208 1208 N-LINKED (POTENTIAL).
 FT CARBOHYD 1233 1233 N-LINKED (POTENTIAL).
 FT CARBOHYD 1237 1237 N-LINKED (POTENTIAL).
 FT CARBOHYD 1332 1332 N-LINKED (POTENTIAL).
 SQ SEQUENCE 1398 AA; 154499 MW; 68E49C4A23C3F824 CRC64;

QY 2 KGLKALIVLVGLV---VGSVAAPAEKKVEQVRNKKNYGLT-----POLFRKIQKLN 54
 DB 4 KGLTVLFTALMLLSVPPVHVSAGTPPVSSSENSTSLPQQVTKVKSQAALNAIKGO 63
 QY 55 PNEESTVIVENHREKETAIVRLMCAKVRVYHIIPAIADL---KYRDLVLISGL- 110
 DB 64 PN---WLIITKREGLEAKTELEKUGAEILDENRVLNKLVKIKPEKVELNLSLE 120
 QY 111 ---TGGKAKLSVURFOEDYKVTYSAELEGLDESAQAQVH--ATYNNNUGDGGITIGII 165
 DB 121 KANLUREVKUS-PPIVEKDVK-TREPSLEPKMNSTWVINALQFTQFEGYDGGSVVAVL 178
 QY 166 DTGIDASH-----PDLQCKVIGWDF-----VNGR----- 190
 DB 179 DTGVDNHPFLSTPDGRKRIENKDKFTDGEFVDTSPFSKVMGTIIITTFQVAGSLT 238
 QY 191 ----- 190
 DB 239 LNESTGLMEYVYVYVSWTIGNITSANGIYHFGLLPERYFDLNFQDQEDFPYVLLVN 298
 QY 191 ----- 194
 DB 299 STGNGYDIAYVDTLOYFTDEVPLOYNYTDVAVFVSYVYVGLNVLAEIDNPGEYAVF 358
 QY 195 ---DRHGHTHVASIAAGTGAAS-----NCKY-----KGMAPG 224
 DB 359 GMDGHHGTHVAGTVAGVDSNDADWLSMYSEMEVESRGLYGHDTYNTTDTVGVAAPG 418
 QY 225 AKLAGIKVADSGSISTILKVERAVONKDKYIKVINLSL-GSSQSSDGTDSLQAV 283
 DB 419 AQIMAIRVLRSDGRGSHWDIIEGMTAA-----THGADVISMGLGNAPYLDGTDPSVAV 474
 QY 284 NNAWDA-GIWCVAAGNSGPNITYVGPAAASKVITVGAVD----- 323
 DB 475 DELTEKYGVVFIAGNNEGPGINIVSGVATKAITVGAAPVINGVYVVSQALGYPDY 534
 QY 324 -----SNDNIAFSRSGPTADGRLKPEVAVGVDIAPRAGSTNSCTPINOYTKAS 375
 DB 535 GPYFYFPAVNIAPFSSRPRIDGEBIKPNVAPGYIY-----SSLPMWICGADF---MS 587
 QY 376 GTSMAHPVSGVAGALIQ-A-HPG---WTPDKVKTALILETA-----DIVAPKEIADIAYA 426
 DB 588 GTSMAHPVSGVALLSGPKPGIYINPDIIKKVLESGATWLEGGDPTGKYTELDOGH 647
 QY 427 GRVWV-----YKAI-----KYDDYAKL----- 443
 DB 648 GLVWYTKSWELKALNTTLPVVDHNAKYSDFAEYLVGVDTIRGLYARNISIPDVEWHI 707
 QY 444 -----TF-----TGVS----- 449
 DB 708 KYVGDTERTFEIYATEPWIKFPVSGVILENNTEFVLRKVDVBSGLEPGLVGRIIID 767
 QY 450 -----ADKGSATHT-EDVSGATFT-----ATLW 473
 DB 768 PTPVIEDEILMTIVPEKFTPENNYTLTWYDINGPQWTHHFTVPECDVLYAMTTW 827

QY 474 DTGSSDIDLVLDPNG-----NEVDYSVYAYVGEKVKYVYNPTAGTWTYKVVSYKGAANY 528
 DB 828 DYG-----LYRPGDNEVFPYQDYLPA-----VSNPMPGMEVLTWTCNEAPLY 872
 QY 529 QVDVSDGSLSGSGGNPNPNPNPTTDTOTFTGSVN-DYWDSDSTFTNNVN----- 582
 DB 873 E-----SGFLVRINGVBITPS-----VWYINRYLDTNTFESEFNITNIY 913
 QY 583 -----SCATKATGDLTF-----DTSYNDLDLVLV 606
 DB 914 APINATLPIGLCTYNASVESVGDGEFFIKGIEVPECTABLKIRIGNPSVPSNLDLVLV 973
 QY 607 DPNGLNLYDRSTSSNSYEHVEYAMPAPGTWTLFVAYS 643
 DB 974 DSKGNLVALDGNPTAREEVVVEYPRPGVYSIVVHGYS 1010

RESULT 4
 ID ELVA_BACAO STANDARD; PRT; 380 AA.
 AC P27593;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Alkaline protease precursor (EC 3.4.21.-).
 OS Bacillus alcalophilus.
 OC Bacillus; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1445;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PB92;
 RX MEDLINE=91282483; PubMed=2059048;
 RA van der Laan J.C., Gerritse G., Mulleners L.J.M., van der Hoek R.A.,
 "Quax H.J.;
 RT "Cloning, characterization, and multiple chromosomal integration of a
 RT Bacillus alkaline protease gene.";
 RL Appl. Environ. Microbiol. 57:901-909(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).
 RC STRAIN=PB92;
 RX MEDLINE=92390330; PubMed=1518788;
 RA van der Laan J.C., Teplyakov A.V., Kelders H., Kalk K.H., Misset O.,
 Mulleners L.J.M., Dijkstra B.W.;
 RT "Crystal structure of the high-alkaline serine protease PB92 from
 RT Bacillus alcalophilus.";
 RL Protein Eng. 5:405-411(1992).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).
 RX MEDLINE=93078250; PubMed=1447775;
 RA Sobek H., Hecht H.-J., Aehle W., Schomburg D.;
 RT "X-ray structure determination and comparison of two crystal forms of
 RT a variant (Asp115Arg) of the alkaline protease from Bacillus
 RT alcalophilus refined at 1.85 Å resolution.";
 RL J. Mol. Biol. 248:108-117(1992).
 RN [4]
 RP STRUCTURE BY NMR OF 112-380.
 RC STRAIN=PB92;
 RX MEDLINE=97277237; PubMed=9115441;
 RA Martin J.R., Mulder F.A., Karimi-Nejad Y., van der Zwan J.,
 Mariani M., Schipper D., Boelens R.;
 RT "The solution structure of serine protease PB92 from Bacillus
 RT alcalophilus presents a rigid fold with a flexible substrate-binding
 RT site.";
 RL Structure 5:521-532(1997).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

Thu Nov 7 10:12:13 2002

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M65086; AAA22212.1; ;
EMBL: A13738; CAA01128.1; ;
PIR: A49778; A49778;
PDB: 1AH2; 15-APR-98.

DR MEROPS: S08.103;
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00139; SUBTILASE_SER; 1.
KW Hydroxylase; Serine protease; zymogen; Signal; 3D-structure.
FT SIGNAL 1 27
FT PROPEP 28 112
FT CHAIN 113 380
FT ACT_SITE 143 143
FT ACT_SITE 173 173
FT ACT_SITE 326 326
FT ACT_SITE 380 AA; 38853 MW; 539EA72771B6682C CRC64;
SQ SEQUENCE 380 AA; 38853 MW; 539EA72771B6682C CRC64;

Query Match 15.7%; Score 537.5; DB 1; Length 380;
Best Local Similarity 33.6%; Pred. No. 1.3e-23;
Matches 147; Conservative 74; Mismatches 140; Indels 77; Gaps 17;

QY 14 LGLVGVSA-----AAPEKKEQVRNVEKNYGLTLPGLFRKIQKLNPEIS 60
DB 5 LGLKIVASTALLISVAFSSSIASAAEEAKEKYLIGFNEQ-----EAVSEFVEQVEANDEVA 59
QY 61 TVIVFENHREKEIAVRVLELNGAKVRYVYTHIPAIADLVKRDLLVLSGLTGKAKLSGV 120
DB 60 IL-----SEEEVEIELL-----HEFETIPVLSVELSPED---VDALELDPF-----I 99
QY 121 RIQEDYKTVYSALBGLDESAQAQVATVYNNLYGDSGITIGITIDGIDASHPDLOGKV 180
DB 100 SYIEDAEVYTIMA--QSPVNGISRVQAPAHNRGLTSGVKVAVLDGTI--STHPDL--NI 154
QY 181 IGWDFVNGRSYPYDDHGHGTHVAVSIAGTGAASNGK--YKGMAGKAGIKVLAGSGS 238
DB 155 RCGASFVPGEPSTQDNGHGHGTHV-----AGTALNNSIGVLAPEAFYAVKVLGASGS 210
QY 239 GSITIKGVWVYDNRKDYGIKYINLSLGSOSSDGTSLSCAVNNWAGIVYCVVAG 298
DB 211 GSVSSIAQGLEWAGN-----GMHVANLSLGSFSPS---ATLEQAVNSATSGVLYVWASG 263
QY 299 NSGPNYTVYGPAAKSVITVGVADSDNDNIASFSSRQPTADGRLEPYVYVPGVDIAPRA 358
DB 264 NSGAG--SISYPARYANAVAGATDNNNRASFQYAGLD-----IVAPGVNVOSTYP 315
QY 359 SGTSMTGPTNDYNYKAGTSMATPHVSGVGLIILQAPHSWTPDKVKYKTALETADIVAPKE 418
DB 316 GST-----YASLNGTSMATPHVGAALVYKQKPSWNSVNIQRLNKLNTATSLGSTN 366
QY 419 IADIYAGKRVNWKAKR 436
DB 367 L-----YSGLVNVAEATR 380

RESULT 5
ID ELYA_BACCS STANDARD; PRT: 380 AA.
AC ELYA_BACCS
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alkaline protease precursor (EC 3.4.21.-)
OS Bacillus clausii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.

NCBI_TaxID=79880;
[1] SEQUENCE FROM N.A.
PC STRAIN-221 / ATCC 21522 / JCM 9139 / DSM 2512;
RX MEDLINE-93043753; PubMed-1368952;
RA Takami H., Kobayashi T., Kobayashi M., Yamamoto M., Nakamura S.,
RA Xono R., Horikoshi K.;
RT "Molecular cloning, nucleotide sequence, and expression of the
ET structural gene for alkaline serine protease from alkaliphilic
ET Bacillus sp. 221.";
RL BioSci. Biotechnol. Biochem. 56:1455-1460(1992).
[2] SEQUENCE OF 112-129.
PC STRAIN-221 / ATCC 21522 / JCM 9139 / DSM 2512;
RA HORIKOSHI K.;
RL (in) Horikoshi K. (eds.);
RL Microorganisms in alkaline environments, pp.187-194, VCH,
RL Weinheim (1991).
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL: S48734; AAC60420.1; ;
EMBL: D13157; BAA02442.1; ;
EMBL: A26817; CAA01836.1; ;
EMBL: A22550; CAA01611.1; ;
DR HSP: P29600; IGCI.
DR MEROPS: S08.103;
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00139; SUBTILASE_SER; 1.
KW Hydroxylase; Serine protease; zymogen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 111
FT CHAIN 112 380
FT ACT_SITE 143 143
FT ACT_SITE 173 173
FT ACT_SITE 326 326
FT ACT_SITE 380 AA; 38826 MW; 5F73ABC6B5B6831 CRC64;
SQ SEQUENCE 380 AA; 38826 MW; 5F73ABC6B5B6831 CRC64;
Query Match 15.7%; Score 537.5; DB 1; Length 380;
Best Local Similarity 33.6%; Pred. No. 1.3e-23;
Matches 147; Conservative 74; Mismatches 140; Indels 77; Gaps 17;

QY 14 LGLVGVSA-----AAPEKKEQVRNVEKNYGLTLPGLFRKIQKLNPEIS 60
DB 5 LGLKIVASTALLISVAFSSSIASAAEEAKEKYLIGFNEQ-----EAVSEFVEQVEANDEVA 59
QY 61 TVIVFENHREKEIAVRVLELNGAKVRYVYTHIPAIADLVKRDLLVLSGLTGKAKLSGV 120
DB 60 IL-----SEEEVEIELL-----HEFETIPVLSVELSPED---VDALELDPF-----I 99
QY 121 RIQEDYKTVYSALBGLDESAQAQVATVYNNLYGDSGITIGITIDGIDASHPDLOGKV 180
DB 100 SYIEDAEVYTIMA--QSPVNGISRVQAPAHNRGLTSGVKVAVLDGTI--STHPDL--NI 154
QY 181 IGWDFVNGRSYPYDDHGHGTHVAVSIAGTGAASNGK--YKGMAGKAGIKVLAGSGS 238
DB 155 RCGASFVPGEPSTQDNGHGHGTHV-----AGTALNNSIGVLAPEAFYAVKVLGASGS 210
QY 239 GSITIKGVWVYDNRKDYGIKYINLSLGSOSSDGTSLSCAVNNWAGIVYCVVAG 298

DR PROSITE: PS00136; SUBTILASE ASP; 1.
DR PROSITE: PS00137; SUBTILASE HIS; 1.
DR PROSITE: PS00138; SUBTILASE SER; 1.
KW Hydrolyase; Sporulation; Serine protease; Zymogen; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 106 POTENTIAL.
FT CHAIN 107 381 SUBTILISIN MAT.
FT ACT_SITE 138 138 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 170 170 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 327 327 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 381 AA; 39507 MW; DAED4B16ED1BA092 CRC64;

Query Match 15.3%; Score 524; DB 1; Length 381;
Best Local Similarity 37.9%; Pred. No. 7.6e-23;
Matches 145; Conservative 53; Mismatches 127; Indels 58; Gaps 15;

OY 61 TVIVFENHREKEIAVRVLELNGAKVRYVYHIIPAIADL---KVRDLVLSGLTGKAKL 117
DB 46 TMSAMSSAKKD---VISEKGGKVKQFKVYVNAATAATLDEKAVKEL-----KKD 91
OY 118 SGVRFIOEDYKTVTSAB-LEGDESAAQVMATYVNNLGYDGSGITIGITIDGIDASHPDL 176
DB 92 PSVAYVEEDH---IAHEYAQSPYVIGISQIKAPALHSGQYTCGMNVKAVIDSIGDSSHPDL 148
OY 177 QGKVIQEDYKTVTSAB-LEGDESAAQVMATYVNNLGYDGSGITIGITIDGIDASHPDL 176
DB 92 PSVAYVEEDH---IAHEYAQSPYVIGISQIKAPALHSGQYTCGMNVKAVIDSIGDSSHPDL 148
OY 234 GADGSGSISTIKGVEMAYDNKDKYGIKVINLSGSSQSDGTDLSQAVNNADAGIVV 293
DB 203 DSTGSGQYSWINGIENWISN---NDVYINNSLG---GPTGSLTKTVVDKAVSSGIVV 255
OY 294 CVAAGNSGP---NTTVGSPAMAKVITVCAVDNDNTASPSRGPATDGLKPKVAVPGV 351
DB 256 AAAAGNEGSGSTSTVGYPAKYFSTIAGVAVNNSSNQPSFSSAGSGLD-----VMAFGV 309
OY 352 DIAPRASGTSMGTPIINDYTKASGTSMATPHVSGVCAILIQHPSTPDVKYKTALETA 411
DB 310 SIQSTLPGGT-----YGAINGTSMATPHVAGAAALILSKHPTWNAQVRDLRESTA 360
OY 412 DIVAPKEIADIAGGRVNYKA 434

Query Match 15.3%; Score 523; DB 1; Length 381;
Best Local Similarity 37.6%; Pred. No. 8.7e-23;
Matches 144; Conservative 54; Mismatches 127; Indels 58; Gaps 15;

OY 61 TVIVFENHREKEIAVRVLELNGAKVRYVYHIIPAIADL---KVRDLVLSGLTGKAKL 117
DB 46 TMSAMSSAKKD---VISEKGGKVKQFKVYVNAATAATLDEKAVKEL-----KKD 91
OY 118 SGVRFIOEDYKTVTSAB-LEGDESAAQVMATYVNNLGYDGSGITIGITIDGIDASHPDL 176
DB 92 PSVAYVEEDH---IAHEYAQSPYVIGISQIKAPALHSGQYTCGMNVKAVIDSIGDSSHPDL 148
OY 177 QGKVIQEDYKTVTSAB-LEGDESAAQVMATYVNNLGYDGSGITIGITIDGIDASHPDL 233
DB 149 --NVRGGASFPVSETNPYQDGSSTHGV---AGTIAALNNSIGVLGVSFASIAVKVL 202
OY 234 GADGSGSISTIKGVEMAYDNKDKYGIKVINLSGSSQSDGTDLSQAVNNADAGIVV 293
DB 203 DSTGSGQYSWINGIENWISN---NDVYINNSLG---GPTGSLTKTVVDKAVSSGIVV 255
OY 294 CVAAGNSGP---NTTVGSPAMAKVITVCAVDNDNTASPSRGPATDGLKPKVAVPGV 351
DB 256 AAAAGNEGSGSTSTVGYPAKYFSTIAGVAVNNSSNQPSFSSAGSGLD-----VMAFGV 309
OY 352 DIAPRASGTSMGTPIINDYTKASGTSMATPHVSGVCAILIQHPSTPDVKYKTALETA 411
DB 310 SIQSTLPGGT-----YGAINGTSMATPHVAGAAALILSKHPTWNAQVRDLRESTA 360
OY 412 DIVAPKEIADIAGGRVNYKA 434

RESULT 8
SUBT_BACSU
ID SUBT_BACSU STANDARD; PRT: 381 AA.
AC P04189.
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Subtilisin E precursor (EC 3.4.21.62).
GN APRE OR APR OR APR OR SPRE.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
CX NCBI_TaxID=1423;
EN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168.
RX MEDLINE=54212198; PubMed=6427178;
RA Stahl M.L., Ferrari E.;
RT "Replacement of the Bacillus subtilis subtilisin structural gene with
an in vitro-derived deletion mutation.";
RL J. Bacteriol. 159:411-418(1984).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX STRAIN=168;
RX MEDLINE=99030466; PubMed=9811547;
RA Jain S.C., Shinde U., Li Y., Inouye M., Berman H.M.;
RT "The crystal structure of an autoprocessed Ser221Cys-subtilisin
E-propeptide complex at 2.0-A resolution.";

Db 361 TYLG-----NSFYKGLINQAA 379

RESULT 9

SUBT_BACSA

ID SUBT_BACSA STANDARD; PRT; 381 AA.
AC P00783;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Subtilisin amylosaccharitic precursor (EC 3.4.21.62).
GN APR.

OS Bacillus subtilis var. amylosacchariticus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1483;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=8908194; PubMed=3139550;

RA Yoshimoto T., Oyama H., Honda T., Tone H., Takeshita T.,

RT Kamiyama T., Tsutsu O.,

RL "Cloning and expression of subtilisin amylosacchariticus gene";

RN J. Biol. Chem. 247:5602-5618(1972).

RN (2)

RP PARTIAL SEQUENCE.

RX MEDLINE=7226687; PubMed=4560201;

RA Markland F.S., Kurihara M., Smith E.L.,

RT "Subtilisin Amylosacchariticus. II. Isolation and sequence of the

RL tryptic and cyanogen bromide peptides";

RN J. Biol. Chem. 247:5602-5618(1972).

RN (3)

RP SEQUENCE OF 107-381.

RX MEDLINE=7226688; PubMed=5055784;

RA Kurihara M., Markland F.S., Smith E.L.,

RT "Subtilisin Amylosacchariticus. 3. Isolation and sequence of the

RL chymotryptic peptides and the complete amino acid sequence";

RN J. Biol. Chem. 247:5619-5631(1972).

RN (4)

CC -1- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE

CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity

CC for peptide bonds, and a preference for a large uncharged residue

CC in pl. Hydrolyses peptide amides.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF

CC SPORELATION, AND MANY MUTATIONS WHICH BLOCK SPORELATION AT EARLY

CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN

CC IS NOT NECESSARY FOR NORMAL SPORELATION.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE

CC SUBTILASE FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL; D00264; BAA00186.1;

CC PIR; A00971; SUBSS.

CC PIR; A14448; A1448.

CC HSSP; P04189; ISCJ.

CC

CC MEROPS; S08.001;

CC InterPro; IPR000209; Peptidase_S8.

CC Pfam; PF00082; Peptidase_S8; 1.

CC PRINTS; PR00723; SUBTILISIN.

CC PROSITE; PS00136; SUBTILASE_ASP; 1.

CC PROSITE; PS00137; SUBTILASE_HIS; 1.

CC PROSITE; PS00138; SUBTILASE_SER; 1.

KW Hydrolase; Sporulation; Serine protease; Zymogen; Signal.

FT SIGNAL 1 30 POTENTIAL.

FT PROPEP 31 106

FT CHAIN 107 381 SUBTILISIN AMYLOSACCHARITICUS.

FT ACT_SITE 138 138 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 170 170 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 327 327 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CONFLICT 191 191 S -> A (IN REF. 2).
FT CONFLICT 365 365 N -> D (IN REF. 2).
SQ SEQUENCE 381 AA; 39467 MW; 2251BADE22B4824F CRC64;

Query Match 15.2%; Score 520; DB 1; Length 381;

Best Local Similarity 37.3%; Pred. No. 1.3e-22;

Matches 143; Conservative 55; Mismatches 127; Indels 58; Gaps 15;

OY 61 TVYFENHREKEIAVRVLELNGAKVRYVYHIPIAIAADL---KVRDLLVLSGLTGGKAKL 117

DB 46 TMSANSSAKKD---VISEKGGKVKQKQKVVNAAATLDEKAVKEL-----KKD 91

OY 118 SGVRPEIDYKVTVSAB-LEGLDESAAQVWATYVWNLGVDSGITIGIIGIDIGIDASHPDL 176

DB 92 PSVAIVEEDH---IAHEYAQSVYVIGISOIKAPALHSGQVTVGSNVKAVYVDSIGDSSHPDL 148

OY 177 OGKVLGVGVNGRSYPYD-HGCTHVASIAGTGAASNGK--YKMAPGAKLAGIKVL 233

DB 149 --NVRGASVPSETPYDQSSGCTH---AGTAAALNNSIGVLGSPSASLYAVKVL 202

OY 234 GADSGSITIIKGVNAVVDNKKYIKVNLGLSGSSQSDGTDLSQAVNNAMDAGIVV 293

DB 203 DSTSGGYSWIINGIEMWISN---NMDVINNSLG---GPSGSTALKTVDKAVSSGIVV 255

OY 294 CVAAGNSGP--NTYTVGSPAASKVITVGAVDNDNIASFSSRGPTADGRLKPEYVAGV 351

DB 256 AAAAGNEGSSGSSVTGVYPAKVPSTIAGVAVNSNORASFSSAGSELD-----VAPGV 309

OY 352 DIAPRASGTSNGTPINDYYTKASGTSKATPHVSGVGLIIOAHPSWTDPKVTALIEFA 411

DB 310 SIQSTLPGT-----YGAINGTSKATPHVAGAAALILSKHPTWTNAQVRDRLESTA 360

OY 412 DIVAPKEIADIYAGRVNYKA 434

DB 361 TYLG-----NSFYKGLINQAA 379

RESULT 10

SUBT_BACST

ID SUBT_BACST STANDARD; PRT; 381 AA.

AC P29142;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Subtilisin J precursor (EC 3.4.21.62).

GN APRJ.

OS Bacillus stearothermophilus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Geobacillus.

OX NCBI_TaxID=1482;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=NCIMB 10278 / KCTC 1823;

RX MEDLINE=92231938; PubMed=1587435;

RA Jiang J.S., Kang D.O., Chun M.J., Byun S.M.;

RT "Molecular cloning of a subtilisin J gene from Bacillus

RI stearothermophilus and its expression in Bacillus subtilis.";

RL Biochem. Biophys. Res. Commun. 184:277-282(1992).

CC -1- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE

CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity

CC for peptide bonds, and a preference for a large uncharged residue

CC in pl. Hydrolyses peptide amides.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF

CC SPORELATION, AND MANY MUTATIONS WHICH BLOCK SPORELATION AT EARLY

CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN

CC IS NOT NECESSARY FOR NORMAL SPORELATION.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE

CC SUBTILASE FAMILY.

CC

RESULT 11	ELYA_BACHD	STANDARD:	PRT:	361 AA.
ID	ELYA_BACHD			
AC	P41363: Q53294; Q9KEJ7;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	15-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Thermotomastix alkaline protease precursor (EC 3.4.21.-).			
GN	BH0855.			
OS	Bacillus halodurans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			

FT	SIGNAL	25	93	TERMOSTABLE ALKALINE PROTEASE.
FT	PROPEP	94	361	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CHAIN	94	124	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	124	154	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	154	307	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	307	38	E -> D (IN STRAIN AH-101).
FT	VARIANT	38	70	KK -> EE (IN STRAIN AH-101).
FT	VARIANT	69	77	KUKK -> ELGN (IN STRAIN AH-101).
FT	VARIANT	74	86	K -> E (IN STRAIN AH-101).
FT	VARIANT	86	104	N -> S (IN STRAIN AH-101).
FT	VARIANT	104	292	N -> Y (IN STRAIN AH-101).
FT	VARIANT	292	345	PS -> SN (IN STRAIN AH-101).
FT	VARIANT	345	377	PS -> SN (IN STRAIN AH-101).
FT	SEQUENCE	361	38116	MM; 24BF004F9E3E8474 CRG64;
SO	SEQUENCE	361	AA;	

Query Match	Score	DB 1;	Length
Best Local Similarity	14.88	508;	361;
Best Local Similarity	33.18	Pred. NO. 5.7e-22;	

552 PKWSVEQIKAAIMNTA--VTLKSDGEVPHNAQAGSARIMNAIKADSL----- 599
 QY 451 DKGSATHTDVSATFTATLYWDGSSDIDLYDPNGNEVDYSYATYGFVKGYNYP 510
 DB 600 -----VSGSYSGTFLKENGNEKNETFTIENQSSIRKSYLEYFNGSGI--S 647
 QY 511 TAGTWYKVVYK- GAANYQVVDVSGSUSGSGGNPNPNPFTTDTOTFTGVS-- 567
 DB 648 TSGTSRVVIPAHOTGKATKVK-----NTRKTKAGTFTGTVIV 686
 QY 568 -----NDYWDTSDFTMVNSGATKITGDLTFD-----SYNDLDLYL 605
 DB 687 REGKTVAKVPTILLIVKEPDY---PRVTSVSVSEGS--VQCTYQIETYLPAAGAEELAFV 741
 QY 606 YDPNGNLUVRS-----TSSNSYHVEYAN-----PAPGTWTELVIA 641
 DB 742 YDSNLDPAGQAGYKQNDKGQYFDWDGTINGGKLPK-GEYULLAYA 788
 RESULT 13
 SUBT_BACPU STANDARD: PRT: 275 AA.
 ID SUBT_BACPU STANDARD: PRT: 275 AA.
 AC P07518:
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Subtilisin (EC 3.4.21.62) (Alkaline mesentericopeptidase).
 GN APR.
 OS Bacillus pumilus (Bacillus mesentericus).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1408;
 RN [1]
 RP SEQUENCE.
 RA Svendsen L., Genov N., Idakieva K.;
 RT "Complete amino acid sequence of alkaline mesentericopeptidase: a
 RT subtilisin isolated from a strain of Bacillus mesentericus.";
 RL FEBS Lett. 196:228-232(1986).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=9217311; PubMed=1793542.
 RA Dauter Z., Betzel C., Genov N., Pison N., Wilson K.S.;
 RA "Complex between the subtilisin from a mesophilic bacterium and the
 RT leech inhibitor eglin-C";
 RL Acta Crystallogr. B 47:707-730(1991).
 CC -1- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
 CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
 CC for peptide bonds, and a preference for a large uncharged residue
 CC in pI hydrolyses peptide amides.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
 CC SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY
 CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
 CC IS NOT NECESSARY FOR NORMAL SPOULATION.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC PIR: A23624; A23624.
 DR PDB: 1MEE; 15-JAN-93.
 DR MEROPS: S08_002;
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Sporulation; Serine protease; 3D-structure.
 FT ACT_SITE 32 32 BY SIMILARITY.
 FT ACT_SITE 64 64 BY SIMILARITY.
 FT ACT_SITE 221 221 BY SIMILARITY.
 FT HELIX 11 12 BY SIMILARITY.
 FT TURN 11 12
 FT HELIX 11 19

20 20
 FT TURN 24 25
 FT TURN 27 32
 FT TURN 37 38
 FT TURN 40 41
 FT TURN 44 49
 FT STRAND 51 51
 FT TURN 52 53
 FT STRAND 54 54
 FT TURN 57 58
 FT TURN 61 62
 FT TURN 65 73
 FT HELIX 86 87
 FT TURN 89 94
 FT STRAND 98 99
 FT TURN 101 101
 FT STRAND 104 116
 FT TURN 117 118
 FT TURN 121 124
 FT STRAND 126 128
 FT STRAND 133 144
 FT TURN 145 146
 FT STRAND 148 152
 FT TURN 159 159
 FT STRAND 160 161
 FT TURN 162 162
 FT STRAND 167 167
 FT TURN 168 170
 FT TURN 172 173
 FT STRAND 175 180
 FT TURN 182 183
 FT STRAND 186 186
 FT TURN 188 189
 FT TURN 194 195
 FT STRAND 198 201
 FT TURN 205 209
 FT TURN 210 212
 FT STRAND 213 217
 FT TURN 220 237
 FT TURN 239 240
 FT TURN 243 252
 FT STRAND 255 255
 FT TURN 260 263
 FT TURN 264 265
 FT STRAND 267 267
 FT TURN 270 273
 FT TURN 274 274
 SQ SEQUENCE 275 AA; 27655 MW; 33BCAB897DBA4170A CRC64;
 Query Match 14.6%; Score 500.5; DB 1; Length 275;
 Best Local Similarity 42.8%; Pred. No. 1, 1e-21;
 Matches 127; Conservative 35; Mismatches 98; Indels 37; Gaps 10;
 QY 143 AOVNATYVNLAYDGSCTTIGITDIDASHDHPDLOGKVIHVDYNGRSPYDD-HGHGT 201
 DB 9 SOIKAPALHSQGYTGSNTKVAIDSGIDSSHPL--NVRGASFPVSETPHYQDSSHGT 66
 QY 202 HVASIAAGTGAASNCK--YKGMAPGAKIAGIKVLGADGSGSISTIIKGYEMAVDNKDYG 259
 DB 67 HV-----AGTIALNNSIGVLGAPSSALYAVKVLDTSGTSGOYSWIINGIEWALSIN---N 118
 QY 260 TKVINLSIGSSQSDGDTSLSQAYNNANDAGIVVCVAAAGNSGP--NTYTVGSFAAASKVI 317
 DB 119 KDVINNLSG---GPTGSTATKTVVDKAVSSGIVAAAGNEGSSGSTSTVGYPAKYPSTI 175
 QY 318 TVGAVDSNDNIASFSSRGFTADCKLKEPVVAPGVDIAPRSGTSMGTPINDIYTKASGT 377
 DB 176 AVGAVNSANQRASFSSAGSELD-----VMAPGVSIQSTLPGGT-----YGAINGT 220
 QY 378 SMATPHVSGVAILLAQHPSTPKVKLTALIEDIVAPKEIADIAYGAGRVNYKA 434
 DB 221 SMATPHVGAALILSKHFTWNAQVRDLLESTATVLG----SSFYKGLINYOAA 273

Thu Nov 7 10:12:13 2002

```

FT TURN 166 167
FT STRAND 168 174
FT TURN 176 177
FT STRAND 178 180
FT TURN 182 183
FT STRAND 184 189
FT TURN 192 195
FT STRAND 197 203
FT TURN 204 206
FT STRAND 207 211
FT TURN 214 231
FT STRAND 233 234
FT TURN 237 246
FT STRAND 247 247
FT TURN 249 249
FT STRAND 254 257
FT TURN 258 259
FT STRAND 261 261
FT TURN 264 267
FT STRAND 268 268
SQ SEQUENCE 269 AA; 26823 MW; B9AFF1A6A9E2676B CRC64;

Query Match 14.4%; Score 493.5; DB 1; Length 269;
Best Local Similarity 40.9%; Pred. No. 2.6e-21;
Matches 121; Conservative 46; Mismatches 92; Indels 37; Caps 10;

Qy 143 AQMATYVYNLGYDGGSGITIGIDTGDASHPDILQGVIGVWDFVNGRSYFYDDRHGHTH 202
Db 9 SRVQAPAHNKLIGSGVKVAVLDTGI-STHPDL--NIRGGASFVPEFSTQDGNHGHTH 65
Qy 203 VASIAAGTGAASNK--YKMAPGAKIAGIKVLGADGSGSISTIIKGVEMAVDNKDKYGI 260
Db 66 V----AGTIAALNNSIGVLGVAPSLEYAVKVLGADGGRGAISSIAOGLERAGNN----GM 117
Qy 261 KVINLSGSSQSDGDTLSQAVNNAMDAGIYVCYAAGNSGPNITYYVGSFAAASKVITVG 320
Db 118 HVANLSGSPSPS---ATLEQAVNSATSRGLVVAASGNSGAS--SISYPARYANAMAVG 172
Qy 321 AVDSNDNIASFSSRGPTADGRLKPEVYAPCVDIIAPRASGTSMTPTINDYITKASGTSMA 380
Db 173 ATQNNRNASFSQYAGLD-----IVAGVNVQSTYFGST-----YASLNGTSM 217
Qy 381 TPHSVGVGALLIQAHPSWTPDKVKTALIFTADIYAPKEIDIAYGAGRVNVYKAIK 436
Db 218 TPHVAGAAALVKQKPSWSNVQIRNHLKNTATSLGSTNL-----YGSGLVNAEAATR 269

Search completed: October 31, 2002, 13:30:12
Job time : 15 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:27:48 : Search time 17 Seconds
(without alignments)
3724.875 Million cell updates/sec

Title: US-09-841-553-5

Perfect score: 3428
Sequence: 1 MGLKALILYLVLGLVWGS.....YAVSTYGMADYOLKAVVYVG 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pr1.*
2: pr2.*
3: pr3.*
4: pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	688	20.1	442	2	A69587 intracellular alk
2	667.5	19.5	444	2	B83891 intracellular alka
3	555	16.2	382	1	SUBSN subtilisin (EC 3.4
4	554	16.2	379	1	SUBSCL subtilisin (EC 3.4
5	544	15.9	382	2	I39780 subtilisin (EC 3.4
6	540.5	15.8	1398	2	T28159 subtilisin (EC 3.4
7	537.5	15.7	380	2	A49778 pycolysin (EC 3.4
8	535	15.6	274	1	SUBSD high-alkaline seri
9	524	15.3	381	2	JH0778 subtilisin (EC 3.4
10	523	15.3	275	2	JH0778 subtilisin (EC 3.4
11	523	15.3	381	1	SUBSI subtilisin (EC 3.4
12	520	15.2	381	1	SUBSS subtilisin (EC 3.4
13	520	15.2	381	2	JQ1487 subtilisin (EC 3.4
14	508	14.8	361	2	A48373 subtilisin (EC 3.4
15	508	14.8	361	2	G83756 high-alkaline seri
16	505.5	14.7	806	2	A1341 subtilisin-type al
17	498.5	14.5	272	2	A23624 microbial serine p
18	496.5	14.5	384	2	JC4802 subtilisin (EC 3.4
19	496.5	14.5	757	2	C84120 subtilisin (EC 3.4
20	496	14.5	374	2	I39781 subtilisin-type al
21	491.5	14.3	401	2	I39974 microbial serine p
22	489.5	14.3	799	1	SUNYTV alkaline proteinase
23	488	14.2	759	2	G83753 serine proteinase
24	486.5	14.2	378	2	A33973 thermolysin (EC 3.4
25	460	13.4	715	2	JC4908 high-alkaline seri
26	452.5	13.2	1433	1	A36734 alkaline serine pr
27	451.5	13.2	372	2	D83735 bacillopeptidase F
28	449	13.1	627	2	D75393 subtilisin-type al
29	446.5	13.0	535	2	B82358 serine proteinase, alkaline serine pr

ALIGNMENTS

RESULT 1

A69587

intracellular alkaline serine proteinase aprX - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999

R:Kunat, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bar

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Gallizzi, A.; Gal

lechi, J.; Harwood, C.R.; Hensaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hulio, M

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl

A:Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Skovska, A.; Sa

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil

A:Reference number: A69580; MUID:98044033

A:Accession: A69587

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-442 <KUN>

A:Cross-references: GB:Z99113; GB:A0009126; NID:92634090; PIDN:CAB13610.1; PID:e11833

A:Experimental source: strain 168

C:Genetics:

A:Gene: aprX

C:Superfamily: subtilisin homology

F:146-398/Domain: subtilisin homology <SBT>

Query Match

Best Local Similarity 20.1%; Score 688; DB 2; Length 442;

Matches 167; Conservative 78; Mismatches 132; Indels 44; Gaps 14;

QY	43	TPGLRKI--OKLNPEEISTVIVFENHREK---EIAVRVLEL-NGAKRVYVHIPTA 96
DB	35	TPCFLIKFPETKLQNKMSVIEFEEGCHETGFMAGEVLQKEKSKLSRENKNCSS 94
QY	97	ADLKRDLVLISGLTGGKAKLSGVRFQEDYKVTYSALXGLDSEAAQVM-ATYVNLG- 154
DB	95	AE-----VTPALLUSLUSECNR-----KYUNREVKALLDTATYASHAKEVVRNQ 142
QY	155	-YCGSGITITGIDTIDASHDLOGKVIQYGVDFGRSPYDDHGHGTHVASTAAGTGA 213
DB	143	TLTGKGVTVAVYDTGI-YPHPDLEGRIGFADKNVQKTEPYDDNGHCHTCAGDVASS 201
QY	214	SNKYGKMAPGAKLAGIKVAGDSCSGSISTIKGVEMAV----DNKDKYGIKVINLS 269
DB	202	SSQYRGPAPEANLIGVYLNKQSGSTLADIIEGVEMCIQYNEDNPDE-PIDIMSLS 260

GenCore version 5.1.3

Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:27:48 : Search time 17 Seconds
(without alignments)
3724.875 Million cell updates/sec

Title: US-09-841-553-5

Perfect score: 3428
Sequence: 1 MGLKALILYLVLGLVWGS.....YAVSTYGMADYOLKAVVYVG 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pr1.*
2: pr2.*
3: pr3.*
4: pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	688	20.1	442	2	A69587 intracellular alk
2	667.5	19.5	444	2	B83891 intracellular alka
3	555	16.2	382	1	SUBSN subtilisin (EC 3.4
4	554	16.2	379	1	SUBSCL subtilisin (EC 3.4
5	544	15.9	382	2	I39780 subtilisin (EC 3.4
6	540.5	15.8	1398	2	T28159 subtilisin (EC 3.4
7	537.5	15.7	380	2	A49778 pycolysin (EC 3.4
8	535	15.6	274	1	SUBSD high-alkaline seri
9	524	15.3	381	2	JH0778 subtilisin (EC 3.4
10	523	15.3	275	2	JH0778 subtilisin (EC 3.4
11	523	15.3	381	1	SUBSI subtilisin (EC 3.4
12	520	15.2	381	1	SUBSS subtilisin (EC 3.4
13	520	15.2	381	2	JQ1487 subtilisin (EC 3.4
14	508	14.8	361	2	A48373 subtilisin (EC 3.4
15	508	14.8	361	2	G83756 high-alkaline seri
16	505.5	14.7	806	2	A1341 subtilisin-type al
17	498.5	14.5	272	2	A23624 microbial serine p
18	496.5	14.5	384	2	JC4802 subtilisin (EC 3.4
19	496.5	14.5	757	2	C84120 subtilisin (EC 3.4
20	496	14.5	374	2	I39781 subtilisin-type al
21	491.5	14.3	401	2	I39974 microbial serine p
22	489.5	14.3	799	1	SUNYTV alkaline proteinase
23	488	14.2	759	2	G83753 serine proteinase
24	486.5	14.2	378	2	A33973 thermolysin (EC 3.4
25	460	13.4	715	2	JC4908 high-alkaline seri
26	452.5	13.2	1433	1	A36734 alkaline serine pr
27	451.5	13.2	372	2	D83735 bacillopeptidase F
28	449	13.1	627	2	D75393 subtilisin-type al
29	446.5	13.0	535	2	B82358 serine proteinase, alkaline serine pr

QY 270 S---QSDGDTLSOAVNNADAGIVCVAGNAGSPNTYIVGSPAAASKVITVGAVDN- 325
 DB 261 DALRDVHROEDPLVRAVEANSAGIVCVAGNAGSPNTYIVGSPAAASKVITVGAVDN 320
 QY 326 -----DNIAFSRSGPTADGRLEKREVPVAGVDIAPRAGT-----SMGTPINDVYIKA 374
 DB 321 TASSDDDTVASFSRSGPTVYGEKEDILAPGVNIISURSPSYIDKLOKSRVCSQYFTM 380
 QY 375 SGTSMATPHVSGVGLIQAHPSTWPKVKTALLETADIVAKETADTADAGRVNYKA 434
 DB 381 SGTSMATPICAGTAAALIQONPLTDPVKELKNGTD--KMRDEDPNIYGAGAVNAENS 438
 QY 435 I 435
 DB 439 V 439
 RESULT 2
 B83891
 Intracellular alkaline serine proteinase aprX [imported] - Bacillus halodurans (strain
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: B83891
 R:Takami, H.; Nakagane, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masul, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: B83891
 A:Status: preliminary
 A:Molecule type: DNA
 A:Cross-references: GB:AP001513; GB:BA000004; NID:910174345; PIDN:BA055649.1; GSPDB:CN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: aprX

Query Match 19.5%; Score 667.5; DB 2; Length 444;
 Best Local Similarity 36.7%; Pred. No. 2.3e-29;
 Matches 156; Conservative 80; Mismatches 122; Indels 67; Gaps 12;
 QY 35 VERNYGLLPGLFRKIQKNEDISTIVFENHREKEIYRVLELMGAKRVYVHILPA 94
 DB 57 VELNQCFOQG-----SMTVEHVLKHKR-----COLRHSKIRC 92
 QY 95 IAAIDL---KYRDLVWISLGTGKAKLGGVRFIOEDYKVTVSAAELG-LDESAQVMATYV 150
 DB 93 VSANVTPESLQENLVCKDIR-----KIYLNREYHALLDLTAVESAQAPEV 136
 QY 151 WNLG--YDGSITIGITDGTGASHPDLOGKVIQVDFVNGRSYFYDDHCHGTHVASIAA 208
 DB 137 IRNGETITGKDVITIAVDGTI-YPHDELEGRKAFVDFVNDREPPYDDNGHGHGTHCAGDAA 195
 QY 209 GTGAASNGYKYNAPCAKLAGIKVLGADGSGSITIKGYENAVDNKDY---GIKVLNL 265
 DB 195 GNGASSDQVGRGPAPEANVIGYKLVNKGMSLESINQGVENCIOYNEHPDDPIHISM 255
 QY 266 SLGSS---QSSDGTSLQAVNANWDAGIVVCAAGNSGPNNTYIVGSPAAASKVITVGA 322
 DB 256 SLGGQALPYENEQEDPMRIVEEAWNAGITVCVAAGNSGPDQAQTASPGVSEKVIITVGAL 315
 QY 323 D-----SNDNTASFSSRGPTADGRLEKREVPVAGVDIAPRAGT-----SMGTPINDY 370
 DB 316 DDDRTDREDDVAPFSSRGPTIYCKPKPOLAPGVNIISURSPSYIDKLOKSRVCSQYFTM 375
 QY 371 YTKASGTSMATPHVSGVGLIQAHPSTWPKVKTALLETADIVAKETADTADAGRVNYKA 430
 DB 376 YTMNSGTSMATPHVSGVGLIQAHPSTWPKVKTALLETADIVAKETADTADAGRVNYKA 433
 QY 431 YTKAI 435
 DB 434 AEGAI 438

RESULT 3
 SUBSIN
 N:Subtilisin (EC 3.4.21.62) BPN' precursor - Bacillus amyloliquefaciens
 N:Alternate names: subtilisin Novo
 C:Species: Bacillus amyloliquefaciens
 C:Date: 24-Apr-1984 #sequence_revision 28-Aug-1985 #text_change 21-Jul-2000
 C:Accession: B25415; A93495; T44584; A92033; A00970
 R:Vanantol, N.; Thompson, L.D.; Rhodes, C.; Banner, C.; Negle, J.; Filipula, D.
 J. Bacteriol. 159, 811-819, 1984
 A:Title: Genes for alkaline protease and neutral protease from Bacillus amyloliquefac
 A:Reference number: A25415; MUID:85006739
 A:Accession: B25415
 A:Molecule type: DNA
 A:Residues: 1-382 <VAS>
 A:Cross-references: GB:K02496; NID:9142525; PIDN:AA05345.1; PID:9142526
 A:Experimental source: ATCC 23844
 R:Wells, J.A.; Ferrar, E.; Henner, D.J.; Estell, D.A.; Chen, Z.Y.
 Nucleic Acids Res. 11, 7911-7925, 1983
 A:Title: Cloning, sequencing, and secretion of Bacillus amyloliquefaciens subtilisin
 A:Reference number: A93495; MUID:84069812
 A:Accession: A93495
 A:Molecule type: DNA
 A:Residues: 1-382 <WEL>
 A:Accession: T44584
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-382 <WEL>
 A:Cross-references: EMBL:A00165; NID:939337; PIDN:CAA24990.1; PID:9773560
 J. Biol. Chem. 242, 5198-5211, 1967
 R:Markland, F.S.; Smith, E.L.
 A:Title: Subtilisin BPN'. VII. Isolation of cyanogen bromide peptides and the complet
 A:Reference number: A92033; MUID:68086682
 A:Accession: A92033
 A:Molecule type: protein
 A:Residues: 108-162, 'PN', 165-167, 'D', 169-194, 'SA', 197-204, 'DA', 207-264, 'ST', 267-357, '
 R:Krat, J.
 In The Enzymes, 3rd ed., vol. 3, Boyer, P.D., ed., pp.547-560. Academic Press, New Yor
 A:Title: Subtilisin: X-ray structure.
 A:Reference number: A94443
 A:Contents: annotation; X-ray crystallography, 2.5 angstroms; active site
 C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and m
 not necessary for normal sporulation.
 C:Genetics:
 A:Start codon: GTG
 C:Superfamily: subtilisin; subtilisin homology
 C:Keywords: hydrolase; serine proteinase
 P:1-32/Domain: signal sequence #status predicted <SIG>
 P:33-107/Domain: activation peptide #status predicted <APT>
 F:108-382/Product: subtilisin BPN' #status experimental <MPT>
 F:130-342/Domain: subtilisin homology <SPT>
 F:139,171,328/Active site: Asp, His, Ser #status experimental

Query Match 16.2%; Score 555; DB 1; Length 382;
 Best Local Similarity 35.1%; Pred. No. 2.7e-23;
 Matches 158; Conservative 65; Mismatches 141; Indels 86; Gaps 18;
 QY 1 MGLKALILVLVLGLV-----VGSVAAPAEKKVEQVRNVEKNYGLITPGLFRKIQKNPN 56
 DB 1 MRGKRWISLILFALLIFTMAFGSTSSA---QAAGKSNGEKKY----- 40
 QY 57 EISIVVIFENHREKEIYRVLELMGAKRVYVHILPAIADLK---VRDLVLVSLG 110
 DB 41 -----IVGFKQTMTHSAKKKDVISEKGGKQKQFYVDAASTLNKAYKEL----- 89
 QY 111 TGGKARLSGVRFIOEDYKVTVSAAELGDESAQVMATYVNIILGYSGITIIGIDGID 170
 DB 90 ---KKPDSVAYVEEDH--VAHAYAQSVYPYGSQIKAPALHSOGYTGSKNVKAVIDSGID 143
 QY 171 ASHPDLOGKVIQVDFVNGRSYFYDDHCHGTHVASIAAGTGAASNGK--YKNAPCAKL 227
 DB 144 SSRPDL--KYAGASNVPSETNPFQDNNSHGTHV-----AGTVAALNNSIGVLGVAPSASL 197

QY 228 ACIKVLGADGSGSISIIIGKVEWVDNRKDYGIKVINLSLGSQSSDGTDSLQAVNNAM 287
 Db 198 YAVKVLGADSGSGISIIIGKVEWVDNRKDYGIKVINLSLGSQSSDGTDSLQAVNNAM 250
 QY 288 DAGIYVCAAGNSGP--NTYTVGSPAAASKVITVAVDSNDNIASFSSRGPTADGRKPE 345
 Db 251 ASGVVVAAGNEGTSSTGVYPKYPSVIAVAVDSNDNIASFSSRGPTADGRKPE 304
 QY 346 VYAPVDIAPRASGTSMTPTINDYTKASGTSMTAPHSVGVGALILQAHPSWTPDKVKT 405
 Db 305 VYAPVYSI-----QSTLPGNRYGAYNGTSMASPHVAGAAALILSKHPNNTNQVRS 355
 QY 406 ALIETADIVAPKEAD-TAYGAGRVNYKA 434
 Db 356 SLENTT-----TKLGDSPYKGLINVOAA 380

RESULT 4
 SUBSCL
 subtilisin (EC 3.4.21.62) Carlsberg precursor - Bacillus licheniformis
 C:Species: Bacillus licheniformis
 C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 21-Jul-2000
 C:Accession: A24111; A00968
 R:Jacobs, M.; Eliasson, M.; Uhlen, M.; Flock, J.I.
 Nucleic Acids Res. 13, 8913-8926, 1985
 A:Title: Cloning, sequencing and expression of subtilisin from Bacillus licheniformis
 A:Reference number: A24111; MUID:86093688
 A:Accession: A24111
 A:Molecule type: DNA
 A:Residues: 1-379 <JAC>
 A:Cross-references: GB:X03341; NID:9487721; PIDN:CAB56500.1; PID:95921206
 R:Smith, E.L.; DeLange, R.J.; Evans, W.H.; London, M.; Markland, F.S.
 J. Biol. Chem. 243, 2184-2191, 1968
 A:Title: Subtilisin Carlsberg. V. The complete sequence; comparison with subtilisin BPN'
 A:Reference number: A00968; MUID:68234702
 A:Accession: A00968
 A:Molecule type: protein
 A:Residues: 106-206, 'S', 208-232, 'A', 234-261, 'N', 263-264, 'S', 266-315, 'N', 317-379 <SMI>
 C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and many
 not necessary for normal sporulation.
 C:Superfamily: subtilisin; subtilisin homology
 C:Keywords: extracellular protein; signal sequence; status predicted <SIC>
 F:1-29/Domain: signal sequence status predicted <SIC>
 F:30-105/Domain: propeptide status predicted <APT>
 F:106-379/Product: subtilisin Carlsberg status experimental <MPT>
 F:128-339/Domain: subtilisin homology <SBT>
 F:137-168,325/Active site: Asp, His, Ser #status predicted

Query Match 16.2%; Score 554; DB 1; Length 379;
 Best Local Similarity 34.1%; Pred. No. 3e-23;
 Matches 149; Conservative 65; Mismatches 149; Indels 74; Gaps 14;

QY 3 GLKALILVLGLVGVSAAPKVKQVQVNRVKNVGLLTPLGRKTKOKLNPNEISTV 62
 Db 10 GMLTAFMLVTFMFSDSASAAPAK-----NVEKDT-----I 41
 QY 63 IVFENH-RKREIAVRVLELMGAKVRYVTHIIPATAADLVKRVLLVLSGLTGKAKLSGVR 121
 Db 42 VGFSKVTASYVKDILKESGKQKOPRIINAARAKAKLKEALKEVK-----NDPDA 94
 QY 122 FIQEDYKVTYSABLEGDSAAQVMAATYVNNLGVGSGTIGITGIDASHPDLOGKVT 181
 Db 95 YVEEDH--VAHALAQTPVGIPLKADKQVQKPKCAKVAVLDTGQASHPDLP--NVV 150
 QY 182 GWDFVNGRYPVDDHGCHGTHVASTAGTGAASNGK--YKGMAPAKLAGIKVLGADSG 239
 Db 151 CGASFVAGEAYNTDGNHGHV-----ACTVAALDNTTGVGLVAPSVLSYLVKVLNSSGSG 206
 QY 240 SISTIIGVWVDNRKDYGIKVINLSLGSQSSDGTDSLQAVNNAMVAVVCAAGN 299
 Db 207 TYSIVSGIEMATTN-----GMDVINSLG---GPSGSTATKQAVDNAYGVVVVAAAGN 259

QY 300 SGP--NTYTVGSPAAASKVITVAVDSNDNIASFSSRGPTADGRKPEVAVGVDIAPR 357
 Db 260 SSGSGTNTIGYPRKYDSVITAVGAVDSNRSASFSSVG-----AELEVMAFGAVYSTY 313
 QY 358 ASGTSGTNPINDYTKASGTSMTAPHSVGVGALILQAHPSWTPDKVKTALITADIVAPK 417
 Db 314 PTST-----VATLNGTSMASPHVAGAAALILSKHPNLSAQVNRNLSSTATYLG-- 362
 QY 418 ETADIYAGAGRVNYKA 434
 Db 363 --SSFYTGKGLINVEAA 377

RESULT 5
 I39780
 subtilisin (EC 3.4.21.62) Sendai precursor - Bacillus sp.
 C:Species: Bacillus sp.
 C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 22-Jun-1999
 C:Accession: I39780
 R:Yamagata, Y.; Isshiki, K.; Ichishima, E.
 Enzyme Microb. Technol. 17, 653-663, 1995
 A:Title: Subtilisin Sendai from alkalophilic Bacillus sp.: molecular and enzymatic pr
 A:Reference number: I39780; MUID:95329264
 A:Accession: I39780
 A:Status: preliminary; translated from GB/ENBL/DBBJ
 A:Molecule type: DNA
 A:Residues: 1-382 <RES>
 A:Cross-references: GB:D29688; NID:9959363; PIDN:BAA06157.1; PID:9959564
 C:Genetics:
 A:Start codon: TTG
 C:Superfamily: subtilisin; subtilisin homology
 C:Keywords: hydrolase; serine proteinase
 F:136-342/Domain: subtilisin homology <SBT>

Query Match 15.9%; Score 544; DB 2; Length 382;
 Best Local Similarity 33.6%; Pred. No. 1.1e-22;
 Matches 150; Conservative 69; Mismatches 131; Indels 96; Gaps 17;

QY 10 VTLVGLVGVSAAPKVK-----VQVNRVKNVGLLTPLGRKTKOKLNPNE 57
 Db 14 LLLSISLTATSVSAEOKKQVLIQFNQLOVTFVSSDKQO----- 55
 QY 58 EISTVIVENHREKIAVRVLELMGAKVRYVTHIIPATAADLK---VRDLVYISGLTGCK 114
 Db 56 --SMLSFLAEVNDSEIENELL-----YFEDIPVYVSVELSPEDYKDL----- 95
 QY 115 AKLSGVRFQEDYKVTYSABE--GLDESAQVMAATYVNNLGVGSGTIGITGIDIDAS 172
 Db 96 EKQPSITYIEEDIEVITINQVTPMGI-----TRVQAPTATWTRGYTGVRVAVLDGTI-ST 150
 QY 173 HPDLQKVICGWDFVNGRYPVDDHGCHGTHVASTAGTGAASNGK--YKGMAPAKLAGI 230
 Db 151 HPDL--NIRGGVSVFVGPESYQDGNHGHV-----ACTIAALNNSIGVGVGAPNAELVAV 204
 QY 231 KVLGADGSGSITIKGVENAVDNKDYGIKVINLSLGSQSSDGTDSLQAVNNAMVADG 290
 Db 205 KVLGANGSGSVSSAQGLQMTAQN-----NIHVANLSLGSVP---GSOLEAVNQATNAG 257
 QY 291 IVVCAAGNSGNTYTVGSPAAASKVITVAVDSNDNIASFSSRGPTADGRKPEVAVP 350
 Db 258 VLVYAATNGNSG--TVSYPARVANALAVGATDONNRRASFQYGTGLN-----IVAPG 309
 QY 351 VDIAPRASGTSMTPTINDYTKASGTSMTAPHSVGVGALILQAHPSWTPDKVKTALIT 410
 Db 310 VGI-----QSTYPCNRYASLSGTSMTAPHSVAGVAALVKNKPNPSWNTQIROHLTST 360
 QY 411 ADIVAPKEADIYAGAGRVNYKA 436
 Db 361 ATSLGNSH-----QFSGELVNAEAMTR 382

RESULT 6

28159
DB 708 KYVGTETRTETIATPEWIKPFVSGSVILENTEFTVLRKYVDYEGLEPGYVYGRRIID 767
QY 450 -----ADKSATHT-FDVSATFTV-----ATLYW 473
DB 768 PTPVIEDELNTIVTPKPTENNNTLTWTWDINGPEWTHFFFTVPEGVDLYAMTYW 827
QY 474 DTGSSDIDLVDLPNG-----NEWDSYATAYGPERKGVYNTAGTWTWKVSVKGAANY 528
DB 828 DYG-----LYRPGDMFVFPQDLTPAA-----VSNPMPGNWELVWTGNEAPLY 872
QY 529 QVYVYVDSGLSQSGGPNPNPNPTFTTQTFGTSVN-DYWDTSOTFTMNVN-----582
DB 873 E-----SGFLVRHGVETPS-----VWNRITLOTNTEFSEININYY 913
QY 583 -----SGAKITGDLTF-----DTSYNDLDJLY 606
DB 914 APINATLPIGLGTYNASVESVDGEFEIKGIEVPEGTAELKIRIGNPSVPNSOLDJLY 973
QY 607 DPNGLYDRSTSSNSYHEVYANPAGTWTFLVAYS 643
DB 974 DSKGNLVALDGNPTAEVEVYVPEKGVYSIVHGY 1010

RESULT 7
A49778
high-alkaline serine proteinase (EC 3.4.21.-) precursor - Bacillus alcalophilus (stra
N:Alternate names: subtilisin homolog, high-alkaline
C:Species: Bacillus alcalophilus
C:Date: 13-Mar-1994 #sequence_revision 24-Feb-1995 #text_change 20-Jun-2000
C:Accession: A49778; JCI244
C:van der Laan, J.C.; Gerritse, G.; Mulleners, L.J.S.M.; van der Hoek, R.A.C.; Quax,
Appl. Environ. Microbiol. 57, 901-909, 1991
A:Title: Cloning, characterization, and multiple chromosomal integration of a Bacilli
A:Reference number: A49778; MUID:91282483
A:Accession: A49778
A:Molecule type: DNA
A:Residues: 1-380 <VAN>
A:Cross-references: GB:M65086; NID:9142455; PIDN:AAA22212.1; PID:g142457
A:Experimental source: strain PB92, ATCC 31406
A:Note: amino end of mature protein confirmed by peptide sequencing
R:Takami, H.; Kobayashi, T.; Kobayashi, M.; Yamamoto, M.; Nakamura, S.; Aono, R.; Hor
Biosci. Biotechnol. Biochem. 56, 1455-1460, 1992
A:Title: Molecular cloning, nucleotide sequence, and expression of the structural gen
A:Reference number: JCI244; MUID:93043753
A:Accession: JCI244
A:Molecule type: DNA
A:Residues: 1-195, 'S', 197-380 <TAK>
A:Cross-references: GB:D13157; NID:9216231; PIDN:BAA02442.1; PID:g216232
A:Experimental source: Bacillus alcalophilus Vedder, ATCC 21522 (Bacillus sp. 221)
C:Superfamily: subtilisin; subtilisin homolog
C:Keywords: hydrolase; serine proteinase; zymogen
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-111/Domain: activation peptide #status predicted <PRO>
F:112-380/Product: alkaline serine proteinase #status predicted <MAT>
F:134-340/Domain: subtilisin homolog <SBT>
F:143,173,326/Active site: Asp, His, Ser #status predicted

Query Match 15.7%; Score 537.5; DB 2; Length 380;
Best Local Similarity 33.6%; Pred. No. 2.4e-22;
Matches 147; Conservative 74; Mismatches 140; Indels 77; Gaps 17;
QY 14 LGLVVGVA-----AAPEKVRQVRNVEKNYGLTPTGLERKIOKLNPEIS 60
DB 5 LGLVASTALLISVAFSSIASIAEERKYLGFNEQ-----EAYSEFVQVEANDEVA 59
QY 61 TVTVFENHREKEIARVLELMGAKVRVYVHIIPALADLVKRDLLVLSGLTGKAKLSGV 120
DB 60 IL-----SEEEVEIELL-----HEFETIPVLSVELSPED-----VDALELDA---I 99
QY 121 REIQEDKVTVAELEGIDESAAYQMATYVYNNLYGDSGTIGITGDIDASHPDLOGVK 180
DB 100 SYIEDAEVITMA--QSVPWGISRVQAPAAHNRGLTSGYKVAVDITGI--STHPDL---NI 154

28159
DB 2 KGLKALILVILYGVV---VGSVAAAEKKVEQVRNVEKNYGLLT---PGLFRKIOLN 54
QY 4 KGLTLVFIAILSLVVPVHFVSAGTPPVSSSENSTLILNQOVTVKEVSOAALNRMKQ 63
DB 55 PNEBISTVIVFENHREKEIARVLELMGAKVRVYVHIIPALADLVKRDLLVLSGL- 110
QY 64 PN---MVLIIKTREKLEBAKTELEKLGRAILDENRVNMLLVKIKPEKVELNYISLE 120
DB 111 ---TGKAKLSGVRFTQEDYKVTYSNAELEGIDESAAYQMATYVYNNLYGDSGTIGIT 165
QY 121 KAWLNREKLS-PPIVEKDYK-TREPSLEPKMYNSTWVINALQFOEFGYDGSVVAVL 178
QY 166 DFGLDASH-----PDLQKQKVGWDF-----VNGR-----190
DB 179 DTGVDPNHPLSTIPDGRKRIENKDFDTEGFDVDTSEFSKVVNGLIINTTVQASGLT 238
QY 191 -----190
DB 239 LNESTGLMEYVYVTVYVNVITGNITSANGIYHGLLPERYTDLFDGQDQDFPVLVYN 298
QY 191 -----SYPY-----194
DB 299 STGNGYDIAYVDTLDYDFTDEYPLQYNVTYDVAVESYVYGLNLYLAETDPNGEYVF 358
QY 195 ---DDRGCHTHVASTAAGTGAAS-----NGRY-----KGMAPG 224
DB 359 GMDGHGHTHVAGVYAGVDSNDANDWLSMGEWESFRLYKWDYNTVTTDTVQGVAPG 418
QY 225 AKLAGIKVLGADSGSSTTIKGEVAVNDKDKYGIKVINLSL-GSSQSDGTLSQAV 283
DB 419 AQIMATVRLSRDGSMDIIEGTAA-----THGADVISMGLGNAPLYDGTDPESVAV 474
QY 284 NNAWDA-GTWCVAAAGNSGENTTVGSPAASKVITVGAVD-----323
DB 475 DELTERYGVYFVIAAGNEGPGINIVSGPVATKATVGAAPVINGVYVSQALGYDYY 534
QY 324 -----SNDIASFSRGPADGRLEKPEVYAPGVDTIAPRASGTSMGTINDYTKAS 375
DB 535 GFYFPAYTNVRIAFFSRGPRDGEIKDNVAPCGIY-----SSLPMWIGADE--MS 587
QY 376 GTSMATPHVSGVGLILQA-HPS-----WTDQVKTKALIETA-----DIVAPKEIADJAGA 426
DB 588 GTSMATPHVSGVGLIISGPKPEGIYINPDIIKIKVIESGATWLEGDDPYGQKYTELDG 647
QY 427 GRNVN-----YKAI-----KYDDYAKL-----443
DB 646 GLNVNYSWBLKALNGTTLPLVDHWNADKSYSDFAEYLGVDVIRGLVARNSPIDVEMHI 707
QY 444 -----TF-----TGSV-----449

28159
DB 2 KGLKALILVILYGVV---VGSVAAAEKKVEQVRNVEKNYGLLT---PGLFRKIOLN 54
QY 4 KGLTLVFIAILSLVVPVHFVSAGTPPVSSSENSTLILNQOVTVKEVSOAALNRMKQ 63
DB 55 PNEBISTVIVFENHREKEIARVLELMGAKVRVYVHIIPALADLVKRDLLVLSGL- 110
QY 64 PN---MVLIIKTREKLEBAKTELEKLGRAILDENRVNMLLVKIKPEKVELNYISLE 120
DB 111 ---TGKAKLSGVRFTQEDYKVTYSNAELEGIDESAAYQMATYVYNNLYGDSGTIGIT 165
QY 121 KAWLNREKLS-PPIVEKDYK-TREPSLEPKMYNSTWVINALQFOEFGYDGSVVAVL 178
QY 166 DFGLDASH-----PDLQKQKVGWDF-----VNGR-----190
DB 179 DTGVDPNHPLSTIPDGRKRIENKDFDTEGFDVDTSEFSKVVNGLIINTTVQASGLT 238
QY 191 -----190
DB 239 LNESTGLMEYVYVTVYVNVITGNITSANGIYHGLLPERYTDLFDGQDQDFPVLVYN 298
QY 191 -----SYPY-----194
DB 299 STGNGYDIAYVDTLDYDFTDEYPLQYNVTYDVAVESYVYGLNLYLAETDPNGEYVF 358
QY 195 ---DDRGCHTHVASTAAGTGAAS-----NGRY-----KGMAPG 224
DB 359 GMDGHGHTHVAGVYAGVDSNDANDWLSMGEWESFRLYKWDYNTVTTDTVQGVAPG 418
QY 225 AKLAGIKVLGADSGSSTTIKGEVAVNDKDKYGIKVINLSL-GSSQSDGTLSQAV 283
DB 419 AQIMATVRLSRDGSMDIIEGTAA-----THGADVISMGLGNAPLYDGTDPESVAV 474
QY 284 NNAWDA-GTWCVAAAGNSGENTTVGSPAASKVITVGAVD-----323
DB 475 DELTERYGVYFVIAAGNEGPGINIVSGPVATKATVGAAPVINGVYVSQALGYDYY 534
QY 324 -----SNDIASFSRGPADGRLEKPEVYAPGVDTIAPRASGTSMGTINDYTKAS 375
DB 535 GFYFPAYTNVRIAFFSRGPRDGEIKDNVAPCGIY-----SSLPMWIGADE--MS 587
QY 376 GTSMATPHVSGVGLILQA-HPS-----WTDQVKTKALIETA-----DIVAPKEIADJAGA 426
DB 588 GTSMATPHVSGVGLIISGPKPEGIYINPDIIKIKVIESGATWLEGDDPYGQKYTELDG 647
QY 427 GRNVN-----YKAI-----KYDDYAKL-----443
DB 646 GLNVNYSWBLKALNGTTLPLVDHWNADKSYSDFAEYLGVDVIRGLVARNSPIDVEMHI 707
QY 444 -----TF-----TGSV-----449

28159
DB 2 KGLKALILVILYGVV---VGSVAAAEKKVEQVRNVEKNYGLLT---PGLFRKIOLN 54
QY 4 KGLTLVFIAILSLVVPVHFVSAGTPPVSSSENSTLILNQOVTVKEVSOAALNRMKQ 63
DB 55 PNEBISTVIVFENHREKEIARVLELMGAKVRVYVHIIPALADLVKRDLLVLSGL- 110
QY 64 PN---MVLIIKTREKLEBAKTELEKLGRAILDENRVNMLLVKIKPEKVELNYISLE 120
DB 111 ---TGKAKLSGVRFTQEDYKVTYSNAELEGIDESAAYQMATYVYNNLYGDSGTIGIT 165
QY 121 KAWLNREKLS-PPIVEKDYK-TREPSLEPKMYNSTWVINALQFOEFGYDGSVVAVL 178
QY 166 DFGLDASH-----PDLQKQKVGWDF-----VNGR-----190
DB 179 DTGVDPNHPLSTIPDGRKRIENKDFDTEGFDVDTSEFSKVVNGLIINTTVQASGLT 238
QY 191 -----190
DB 239 LNESTGLMEYVYVTVYVNVITGNITSANGIYHGLLPERYTDLFDGQDQDFPVLVYN 298
QY 191 -----SYPY-----194
DB 299 STGNGYDIAYVDTLDYDFTDEYPLQYNVTYDVAVESYVYGLNLYLAETDPNGEYVF 358
QY 195 ---DDRGCHTHVASTAAGTGAAS-----NGRY-----KGMAPG 224
DB 359 GMDGHGHTHVAGVYAGVDSNDANDWLSMGEWESFRLYKWDYNTVTTDTVQGVAPG 418
QY 225 AKLAGIKVLGADSGSSTTIKGEVAVNDKDKYGIKVINLSL-GSSQSDGTLSQAV 283
DB 419 AQIMATVRLSRDGSMDIIEGTAA-----THGADVISMGLGNAPLYDGTDPESVAV 474
QY 284 NNAWDA-GTWCVAAAGNSGENTTVGSPAASKVITVGAVD-----323
DB 475 DELTERYGVYFVIAAGNEGPGINIVSGPVATKATVGAAPVINGVYVSQALGYDYY 534
QY 324 -----SNDIASFSRGPAD

QY 181 IGWDFVNGRSGYPDDHGHGTHVSIAGTGAASNGK--YKGMAPGAKLAGIKVLGADGS 238
 Db 155 RGASGFVPEPSTODGNGHGTIV-----ACTTAALNNSIGVLGVAPNAELVAVKVLGASGS 210
 QY 239 GSISTIKGVEMAVDNKIKYKIKVNLGSLGSSQSDGTSLSQAVNNANDAGIVVCVAG 298
 Db 211 GSVSSIAQGLEWAGNN---GHNANLSLGSPPS---ATLQAVNNSATSGVLVVAAG 263
 QY 299 NSGPNTVTVGSPAAASKVITVCADVSDNTASPSRGPADGRKLPKPEVAPGVDIAPRA 358
 Db 264 NSGAG--SISYPARYANAVAGATDQNNNRASFSQAGLD-----TVAPGVNVQSTYP 315
 QY 359 SCTSWGTPINDYTKASGTSMATPHVSCVGLALQHPSTPDYKTKALITADIVAPKE 418
 Db 316 CST-----FASLNGSTSMATPHVACAAALVKQKNPSNVQVIRNHNKNTATSLGSTN 366
 QY 419 IADYAGAGRVNYKAIK 436
 Db 367 L-----YCSGLVNAEAATR 380

RESULT 8
 SUBSD
 subtilisin (EC 3.4.21.62) DY - Bacillus subtilis (strain DY)
 N:Alternate names: alkaline serine proteinase
 C:Species: Bacillus subtilis
 A:Variety: strain DY
 C:Date: 20-Sep-1984 #sequence_revision 20-Sep-1984 #text_change 02-Jul-1998
 C:Accession: A00969; S02492
 R:Nedkov, P.; Oberthur, W.; Braunitzer, G.
 Biol. Chem. Hoppe-Seyler 366, 421-430, 1985
 A:Title: Determination of the complete amino-acid sequence of subtilisin DY and its comp
 A:Reference number: A00969; MUID:85279895
 A:Accession: A00969
 A:Molecule type: protein
 A:Residues: 1-274 <NED>
 A:Experimental source: strain DY
 B:Zilova, A.; Kleinschmidt, T.; Nedkov, P.
 Biol. Chem. Hoppe-Seyler 366, 1479-1487, 1987
 A:Title: Reductive alkylation of lysine residues in subtilisin DY.
 A:Reference number: S02492; MUID:88134577
 A:Accession: S02492
 A:Molecule type: protein
 A:Residues: 1-212-17;22-24;27-29;43-45;93-95;123-135-137;140-142;169-171;183-187;221;23
 C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and many
 not necessary for normal sporulation.
 C:Superfamily: subtilisin; subtilisin, homology
 C:Keywords: extracellular protein; hydrolase; serine proteinase
 F:23-234/Domain: subtilisin homology <SBT>
 F:32,63,220/Active site: Asp, His, Ser #status predicted

Query Match 15.6%; Score 535; DB 1; Length 274;
 Best Local Similarity 44.9%; Pred. No. 2,1e-22;
 Matches 132; Conservative 36; Mismatches 90; Indels 36; Gaps 9;

QY 145 VMATVWNLGVDGSGITIGIDIGDASHPOLQKVKYGVDFVNGRSGYPDDHGHGTHVA 204
 Db 11 IRADKYQAGYKGVAVYGIIDGTIAASHTDL--KYVGGASFVSGESYNTDNGHGHV- 67
 QY 205 STAAGTGAASNGK--YKGMAPGAKLAGIKVLGADGSISITIKGVEMAVDNKIKYK 262
 Db 68 ---AGTVAALDNTGVLGAPVNSLVAKVLNLSGSGTYSYAIISGLEWATQN----GLDV 120
 QY 263 INLSLGSQSSDGTDSLQAVNNANDAGIVVCVAGNSGP--NTYTVGSPAAASKVITV 320
 Db 121 INKSLG---GPSGTALKQVAKYASGIVVVAAGNSGSSGSGNQNTIGYPAKVDSTAV 177
 QY 321 AVDSNDNTASPSRGPADGRKLPKPEVAPGVDIAPRASGTSMTGPTINDYTKASGTSM 380
 Db 178 AVDSNKNRASFSSVG-----AELEVMAPGVSVISTYPSNT-----YTSLNGTSM 222
 QY 381 TPHSVGVGALLIQAHPSTPDYKTKALITADIVAPKEINDIAYAGRVNYKA 434

Db 223 SPHVAGAAALILSKYPTLSASQVRNRLSTATNLGD----SFYTGKGLINVEAA 272

RESULT 9
 JH0778
 subtilisin (EC 3.4.21.62) NAT precursor - Bacillus subtilis (strain natto NC2-1)
 N:Alternate names: natto proteinase; nattokinase; nattokinase; subtilisin BSP
 C:Species: Bacillus subtilis
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000
 C:Accession: JH0778; JS0601; JS0317; JC2036
 R:Nakamura, T.; Yamagata, Y.; Ichishima, E.
 Biosci. Biotechnol. Biochem. 56, 1869-1871, 1992
 A:Title: Nucleotide sequence of the subtilisin NAT gene, aprN, of Bacillus subtilis
 A:Reference number: JH0778; MUID:93113095
 A:Molecule type: DNA
 A:Residues: 1-381 <NAK>
 A:Cross-references: GB:D25319; NID:9435439; PIDN:BA04989.1; PID:9435440
 R:Sumi, H.
 Kagaku To Seibutsu 29, 119-123, 1991
 A:Title: Natto kinase and fibrinolysis.
 A:Reference number: JS0601
 A:Accession: JS0601
 A:Molecule type: protein
 A:Residues: 107-381 <SUM>
 R:Sumi, H.; Nakajima, N.
 Nippon Nogei Kagaku Kaishi 65, 1125-1127, 1991
 A:Title: Studies on fibrinolysis enzymes in fermentation food.
 A:Reference number: JS0517
 A:Accession: JS0517
 A:Molecule type: protein
 A:Residues: 107-381 <SU2>
 R:Fujita, M.; Nomura, K.; Hong, K.; Ito, Y.; Asada, A.; Nishimuro, S.
 Biochem. Biophys. Res. Commun. 197, 1340-1347, 1993
 A:Title: Purification and characterization of a strong fibrinolytic enzyme (nattokina
 A:Reference number: JC2036; MUID:94107337
 A:Accession: JC2036
 A:Molecule type: protein
 A:Residues: 107-381 <FUJ>
 C:Genetics:
 A:Gene: aprN
 A:Start codon: GTG
 C:Superfamily: subtilisin; subtilisin homology
 C:Keywords: hydrolase; serine proteinase; zymogen
 F:1-39/Domain: signal sequence #status predicted <SIG>
 F:30-106/Domain: activation peptide #status predicted <PRO>
 F:107-381/Product: subtilisin NAT #status experimental <NAT>
 F:129-341/Domain: subtilisin homology <SBT>
 F:138,139,170,327/Active site: Asp, Ser, His, Ser #status predicted

Query Match 15.3%; Score 524; DB 2; Length 381;
 Best Local Similarity 37.9%; Pred. No. 1.3e-21;
 Matches 145; Conservative 53; Mismatches 127; Indels 58; Gaps 15;

QY 61 TVIVFENHREKEITAVRLLELNGAKVRYVYHITATAADL---KYRDLVTSGLTGCAKL 117
 Db 46 TMSANSAKKD-----VISEGGKAVQKQFYVNAATAATLDEKAYKEL-----KND 91
 QY 118 SGVRFQEDYKVTYSAE--LEGDESAQVMAVYVNLGVDGSGITIGIDIGDASHPOL 176
 Db 92 PSVAYVEEDR---IAEYAOQVPYIGISQIKAPALHSQGYTGSNKKVAVIDSGIDSSHPDL 148
 QY 177 QGKVGWDFVNGRSGYPDD--HGHGTHVASIANGTGAASNGK--YKGMAPGAKLAGIKVL 233
 Db 149 --NVRGASGFVSETPNQDSSSHGHV-----AGTIAALNNSIGVLGAPASLAVAKVL 202
 QY 234 GADGSSISITIKGVEMAVDNKIKYKIKVNLGSLGSSQSDGTSLSQAVNNANDAGIV 293
 Db 203 DSTGSGSQSWIINGIENAIEN-----NMDYNNSLG---GPTGSTALKTVYDKAVSSGIW 255
 QY 294 CVAAGNSGP--NTYTVGSPAAASKVITVCADVSDNTASPSRGPADGRKLPKPEVAPCV 351
 Db 256 AAAAGNCGSGSTSTVGYPAKYPSTTAVGVNNSMQRASFSVSGSELD-----VMAPGV 309

A:Title: The subtilisin E gene of *Bacillus subtilis* is transcribed from a sigma37 promoter
A:Accession number: A26116; MUID:84144862
A:Reference: A26116
A:Molecule type: DNA
A:Residues: 1-155 <NON>
A:Cross-references: GB:K01443; NID:g143665; PIDN:AAA22814.1; PID:g143666
R:Ikemura, H.; Takagi, H.; Inoue, M.
J. Biol. Chem. 262, 7859-7864, 1987
A:Title: Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli*
A:Reference number: I39969; MUID:87222417
A:Accession: I39970
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-156 <IK>
A:Cross-references: GB:M16639; NID:g143521; PIDN:AAA22744.1; PID:g143523
R:Henner, D.J.; Ferrar, E.; Perego, M.; Hoch, J.A.
J. Bacteriol. 170, 296-300, 1988
A:Title: Location of the targets of the hpr-97, sacU32(Hy), and sacQ36(Hy) mutations
A:Reference number: I39778; MUID:88086885
A:Accession: I39778
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-8 <HN>
A:Cross-references: GB:M19125; NID:g143527; PIDN:AAA22245.1; PID:g142528
R:Park, S.
J. Bacteriol. 171, 2657-2665, 1989
A:Title: *Bacillus subtilis* subtilisin gene (aprB) is expressed from a sigma-A (sigma-A') promoter
A:Reference number: I39779; MUID:89213955
A:Accession: I39779
A:Molecule type: DNA
A:Residues: 1-13 <PA>
A:Cross-references: GB:M31060; NID:g142529; PIDN:AAA22246.1; PID:g142530
R:Experimental source: strain W168, substrain PV79
R:Kamai, M.; Hoeeg, J.O.; Kaiser, R.; Shafiq, J.; Razzaki, T.; Zaidi, Z.H.; Joernva
FEMS Lett. 374, 363-366, 1995
A:Title: Isolation, characterization and structure of subtilisin from a thermotable
A:Reference number: S68012; MUID:96089945
A:Accession: S68012
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 113-323 <RAM>
R:Kunat, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningsstein, G.; Kroh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ertling, J.; Fabret, C.; Ferrari,
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H

352 DIIAPRASGTSMTGPINDYYTKASGTSMATPHVSGVALILQAHPSWTPDKVKTALIERA 411
Y :
bb :
310 SIQSITLPGTT-----YGAYNGTSMATPHVAGAAALILSKHPTWTNAQVRORLESTA 360
Y :
412 DIVAPKEIADIAYGAGRNVYKA 434
Y :
bb 361 TILG-----NSFYKGLINVQAR 379
Y :

RESULT 10
CIC1085
subtilisin (EC 3.4.21.62); precursor - Bacillus licheniformis
N;Alternate names: alkaline proteinase
C;Species: Bacillus licheniformis
C;Date: 09-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 31-Mar-1997
C;Accession: JCI085
R;Lei, H.; Hong, Y.; Zhang, Y.-Y.; Shen, T.J.
J. Chinese Biochem. J. 9, 441-447, 1993
A;Title: PCR amplifying Cloning and sequencing of the coding sequence of the alkaline H
A;Reference number: JCI085
A;Accession: JCI085
A:Molecule type: DNA
A;Residues: 1-275 <LE>
N;Note: The translation of the start codon ATG is not given in this paper
C;Superfamily: subtilisin; subtilisin homology
C;Keywords: hydrolase; serine protecinase
F;24-235/Domain: subtilisin homology <SBT>
F;33,64,221/Active site: Asp, His, Ser #status predicted

Query Match 15.3%; Score 523; DB 2; Length 275;
Best Local Similarity 43.5%; Pred. No. 9.7e-22;
Matches 128; Conservative 39; Mismatches 91; Indels 36; Gaps 10;

QY 145 VMATYVNIGYDGSGITGIIDTGIDASHPDLOGKVIGWDFVNGRSYPYDDRHGHGTIVA 204
DB 12 IKADKKVAQGFFGANVKNVAVLDTGIOASHEDL--NVYGGASFVAGEAYNTDGNGHGHIV- 68
QY 205 SIAAGTGAASNGK--YKMARGAKLAGIKVLGADSGSISTTIKGVMAYDNKKYGIKV 262
DB 69 ---AGTVAAIDLNTGVGVAVSVSLAYKVLNLSGSGSYSGIVSGIEWATTN----QMDV 121
QY 263 INLSASSOSSDGDLSQAVNNAMDAGIVCVVAGNSGP--NTYTGVSPAASKVIIVG 320
DB 122 IINSLG---GAGSTANKQVDNAYARGVVVAAAAGNSGSSGNTNTIGYPAKYDSVIAVC 178
QY 321 AVDSNDNIASFSSRGPTADGRLKPEVPVAPVDIIAPRASGTSMTGPIINDYYTKASGTSMA 380
DB 179 AVDSNSNRASFSSVG-----AELEVNPQ-----AGVYSTYPTNTYAT-LNGTSMA 223
QY 381 TPHVSGVALILQAHPSWTPDKVKTALIEADIAPKEIADIAYGAGRNVYKA 434
DB 224 SERVAGAAALILSKHNLSASQVRNRSTATYLG-----SSFYKGLINVEAA 273

RESULT 11
SUBSI
subtilisin (EC 3.4.21.62) E precursor - Bacillus subtilis
N;Alternate names: alkaline proteinase; bacillopeptidase E; extracellular alkaline serin
C;Species: Bacillus subtilis
C;Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 16-Jun-2000
C;Accession: A00972; A36116; I35970; I39779; 568012; H69586
R;Stahl, W.B.; Ferrari, E
J. Bacteriol. 158, 411-418, 1984
A;Title: Replacement of the Bacillus subtilis subtilisin structural gene with an in vitro
A;Reference number: A00972; MUID:84212196
A;Accession: A00972
A:Molecule type: DNA
A;Residues: 1-381 <GB>
A;Cross-references: GB:K01988; NID:g143519; PIDN:AAA22742.1; PID:g143520
A;Experimental source: strain 1168
R;Hong, S.L.; Price, C.W.; Goldfarb, D.S.; Doi, R.H.
Mol. and Acad. Sci. U.S.A. 81, 1184-1188, 1984

C:Superfamily: subtilisin; subtilisin homology
 C:Keywords: extracellular protein; hydrolase; protein digestion; serine proteinase
 F:1-23/Domain: signal sequence status predicted <SIG>
 F:24-106/Domain: activation peptide status predicted <APT>
 F:107-381/Product: subtilisin E status predicted <NPT>
 F:129-341/Domain: subtilisin E status predicted <NPT>
 F:138-170/327/Active site: Asp, His, Ser status predicted

Query Match 15.3%; Score 523; DB 1; Length 381;
 Best Local Similarity 37.6%; Pred. No. 1,5e-21;
 Matches 144; Conservative 54; Mismatches 127; Indels 58; Gaps 15;

Qy 61 TVIVFENHREKEIAVRVLELGMKAKVRYVYHIIPATAADL---KYRDLIVISGLTGKAKL 117
 Db 46 TNSAMSSAKKDD---VISEKGGKVKOFKYNVAAAATLDEKAVKEL-----KKD 91
 Qy 118 SGVRFIOEDYKVTYSAB-LEGLDESAQVNAVYVNLGDSGGTIGITIGIDGASHPDL 176
 Db 92 PSVAYVEEDH---IAHEYAQSPVYISQIKAPALHSQVYGSNVKVAVIDSGIDSSHPDL 148
 Qy 177 QGKVIQWDFVNGRSPYDD-HGHTHVASIAAGTGAASNGK--YKGMAPGAKLAGIKVL 233
 Db 149 --NVRGASFPVSETPYQDSSHGTHV---AGTIALNNSIGVGVSPASLYAVKVL 202
 Qy 234 GADGSSISTTIKGVHADVNDKDKYGIKVINLSGSSOSSDGTSLSQAYNNAMDAIVY 293
 Db 203 DSTGSGQYSHWIINGIEMAIN---NMDVINNSLG---GPTGTALKTVDKAVSSGIWV 255
 Qy 294 CVAAGNSGP--NTYVGSAPAAAKVITVGVADSDNDNIASFSSRGPTADGRLKPEVAVPV 351
 Db 256 AAAAGNEGSSGTSVCPYAKYPTSTIAYCAVNSNQRASFSSAGSELD-----VMAPGV 309
 Qy 352 DIIAPRASGTSMTPIINDYTKASGTSMTATPVSCVGLILQAHPSWTPDKVKTALETA 411
 Db 310 STQSLPGCT-----YGAVNGTSMATPHVAGAAALILSKHPTMTNAQVRDRLESTA 360
 Qy 412 DIVAPKEIADYAGGRVNYKA 434
 Db 361 TYLGL-----NSFYKGLINVOAA 379

RESULT 12

SUBS subtilisin (EC 3.4.21.62) amylosaccharitic precursor - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 24-Apr-1994 #sequence_revision 24-Feb-1995 #text_change 16-Jun-2000
 C:Accession: A41448; A00971; S68013
 R:Yoshimoto, T.; Oyama, H.; Honda, T.; Tone, H.; Takeshita, T.; Kamiyama, T.; Tsuru, D.
 J. Biochem. 103, 1060-1065, 1988
 A:Title: Cloning and expression of subtilisin amylosaccharitic gene.
 A:Reference number: A41448; MUID:89008194
 A:Accession: A41448
 A:Molecule type: DNA
 A:Residues: 1-381 <YOS>
 A:CROSS-references: GB:D00264; NID:9216328; PIDN:BAA00186.1; PID:g912425
 A:Experimental source: var. amylosacchariticus
 R:Kurihara, M.; Markland, F.S.; Smith, E.L.
 J. Biol. Chem. 247, 5619-5631, 1972
 A:Title: Subtilisin amylosacchariticus. III. Isolation and sequence of the chymotryptic
 A:Reference number: A00971; MUID:72266588
 A:Accession: A00971
 A:Molecule type: protein
 A:Residues: 107-112-114; 148-152-155-157-164-170-173-174-178-181-200-205-210-212-219-225;
 A:Experimental source: var. amylosacchariticus
 R:Kamal, M.; Hoeve, J.O.; Kaiser, R.; Shafat, J.; Razzak, T.; Zaidi, Z.H.; Joernvall,
 FEBS Lett. 374, 363-366, 1995
 A:Title: Isolation, characterization and structure of subtilisin from a thermostable Bac
 A:Reference number: S68013; MUID:96069945
 A:Accession: S68013
 A:Molecule type: preliminary
 A:Residues: 107-235, 'T', 237-245, 293-381 <KAM>
 A:Comment: Secretion of subtilisin is associated with the onset of sporulation, and many

not necessary for normal sporulation.

C:Genetics:
 A:Start codon: GTG
 C:Superfamily: subtilisin; subtilisin homology
 C:Keywords: extracellular protein; hydrolase; serine proteinase
 F:107-381/Product: subtilisin E status predicted <NAT>
 F:129-341/Domain: subtilisin homology <SBT>
 F:138,170,327/Active site: Asp, His, Ser status predicted

Query Match 15.2%; Score 520; DB 1; Length 381;
 Best Local Similarity 37.3%; Pred. No. 2.2e-21;
 Matches 143; Conservative 55; Mismatches 127; Indels 58; Gaps 15;

Qy 61 TVIVFENHREKEIAVRVLELGMKAKVRYVYHIIPATAADL---KYRDLIVISGLTGKAKL 117
 Db 46 TNSAMSSAKKDD---VISEKGGKVKOFKYNVAAAATLDEKAVKEL-----KKD 91
 Qy 118 SGVRFIOEDYKVTYSAB-LEGLDESAQVNAVYVNLGDSGGTIGITIGIDGASHPDL 176
 Db 92 PSVAYVEEDH---IAHEYAQSPVYISQIKAPALHSQVYGSNVKVAVIDSGIDSSHPDL 148
 Qy 177 QGKVIQWDFVNGRSPYDD-HGHTHVASIAAGTGAASNGK--YKGMAPGAKLAGIKVL 233
 Db 149 --NVRGASFPVSETPYQDSSHGTHV---AGTIALNNSIGVGVSPASLYAVKVL 202
 Qy 234 GADGSSISTTIKGVHADVNDKDKYGIKVINLSGSSOSSDGTSLSQAYNNAMDAIVY 293
 Db 203 DSTGSGQYSHWIINGIEMAIN---NMDVINNSLG---GPTGTALKTVDKAVSSGIWV 255
 Qy 294 CVAAGNSGP--NTYVGSAPAAAKVITVGVADSDNDNIASFSSRGPTADGRLKPEVAVPV 351
 Db 256 AAAAGNEGSSGTSVCPYAKYPTSTIAYCAVNSNQRASFSSAGSELD-----VMAPGV 309
 Qy 352 DIIAPRASGTSMTPIINDYTKASGTSMTATPVSCVGLILQAHPSWTPDKVKTALETA 411
 Db 310 STQSLPGCT-----YGAVNGTSMATPHVAGAAALILSKHPTMTNAQVRDRLESTA 360
 Qy 412 DIVAPKEIADYAGGRVNYKA 434
 Db 361 TYLGL-----NSFYKGLINVOAA 379

RESULT 13

SUBS subtilisin (EC 3.4.21.62) J precursor - Bacillus stearothermophilus
 C:Species: Bacillus stearothermophilus
 C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 22-Jun-1999
 C:Accession: JQ1487
 R:Jang, J.S.; Kang, D.O.; Chun, M.J.; Byun, S.M.
 Biochem. Biophys. Res. Commun. 184, 277-282, 1992
 A:Title: Molecular cloning of a subtilisin J gene from Bacillus stearothermophilus an
 A:Reference number: JQ1487; MUID:92231938
 A:Accession: JQ1487
 A:Molecule type: DNA
 A:Residues: 1-381 <YAN>
 A:CROSS-references: GB:M64743; NID:g142531; PIDN:AAA2247.1; PID:g142532
 A:Experimental source: strain NCINB10278
 C:Genetics:
 A:Gene: aptJ
 A:Start codon: GTG
 C:Superfamily: subtilisin; subtilisin homology
 C:Keywords: hydrolase; serine proteinase; zymogen
 F:1-29/Domain: signal sequence status predicted <SIG>
 F:30-106/Domain: activation peptide status predicted <ACT>
 F:107-381/Product: subtilisin J status predicted <NAT>
 F:129-341/Domain: subtilisin homology <SBT>
 F:138,170,327/Active site: Asp, His, Ser status predicted
 Query Match 15.2%; Score 520; DB 2; Length 381;
 Best Local Similarity 37.3%; Pred. No. 2.2e-21;
 Matches 143; Conservative 55; Mismatches 127; Indels 58; Gaps 15;

Qy 61 TVIVFENHREKEIAVRVLELGMKAKVRYVYHIIPATAADL---KYRDLIVISGLTGKAKL 117

Db 45 TMSANSSAKKD-----VISEKGGKQKQFYVNAATIDKAVREL-----KKD 91
QY 118 SGVRFIOEDYKVTVAE-LEGLDESAQAQVNAATYVNLGVGSGDITIGIDGIDASHPDL 176
Db 92 PSVAYVEDH---IAHEYAQSPYGISQIKAPALHSQGYTGSNVKVVAVJDSGIDSSHPDL 148
QY 177 QGKVLGWDFYNGRSYDD-HGCTHVASIAAGTGAASNGK--YKGMAPGAKLAGIKVL 233
Db 149 --NVRGASFPVSTNPQOGSGHGV-----AGTIAALNNSIGVLOTYSPASLYAVKVL 202
QY 234 GADGSGSTIIKGVAVDNKDKYIKVINLSGSSQSDGTDLSQAVNNAWDAGIYV 293
Db 203 DSTGSGOYSWIINGIENAIN---NNDVINNSLG---GSPSGTALKTVYDKAVSSGTV 255
QY 294 CVAAGNSCP--NTVYSGPAAASKVITVGAVDSDNDNTASPSRGTADGRKLPKPEVAPGV 351
Db 256 AAGAGNCGSSGSSSTVGPAPKPSITAVGAVNNSNORASFSAGSELD-----VMAPCV 309
QY 352 DIIAPRASGTSNGTPINDYTKASGTSMATPHVSGVALLILOAHPSWTPDKVKTALITETA 411
Db 310 SIQSTLPQGT-----YGAVNGISMATPHVAGVAAALDKKHPITWNAQVDRLESTA 360
QY 412 DIVAPKEIADYAGGRVNYKA 434
Db 361 TVLG-----NSFTYKGLINVQAA 379

RESULT 14
A48373
high-alkaline serine proteinase (EC 3.4.21.-) precursor - Bacillus sp. (strain AH-101)
N:Alternate names: subtilisin-like thermostable alkaline serine proteinase
C:Species: Bacillus sp.
C/Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 28-May-1999
C/Accession: A48373; J50714
R:Takami, H.; Kobayashi, T.; Aono, R.; Horikoshi, K.
Appl. Microbiol. Biotechnol. 38, 101-108, 1992
A/Title: Molecular cloning, nucleotide sequence and expression of the structural gene for subtilisin-like thermostable alkaline serine proteinase from Bacillus sp. strain AH-101
A/Reference number: A48373; MUID:93098926
A/Accession: A48373
A/Molecule type: DNA
A/Residues: 1-361 <TA2>
A/Cross-references: GB:S50880; NID:g261737; PIDN:AAC60421.1; PID:g261738
A/Experimental source: AH-101
A/Note: this sequence is inconsistent with the nucleotide translation
R:Takami, H.; Kobayashi, T.; Aono, R.; Horikoshi, K.
submitted to JIPID, July 1992
A/Description: Molecular cloning, nucleotide sequence and expression of the structural gene for subtilisin-like thermostable alkaline serine proteinase from Bacillus sp. strain AH-101
A/Reference number: J50714
A/Accession: J50714
A/Molecule type: DNA
A/Residues: 34-334, L', 336-361 <TA2>
C/Comment: This alkaliphilic Bacillus homolog to the subtilisins of neutrophilic Bacillus
C/Superfamily: subtilisin; subtilisin homology
C/Keywords: extracellular protein; hydrolase; serine proteinase
F:115-321/Domain: subtilisin homology <SBT>
F:124,154,307/Active site: Asp, His, Ser #status predicted

Query Match 14.8%; Score 508; DB 2; Length 361;
Best Local Similarity 33.7%; Pred. No. 9.3e-21;
Matches 147; Conservative 79; Mismatches 128; Indels 82; Gaps 17;

QY 2 KGLKALILVLGLVGSVAAPEKKEVQVRNVEKNYGLLTPGLFRKIQKLPNEEIST 61
Db 3 QSLKVMVLSVAL-LFMANPAAASEEKEYLIVPEP-----DEVA 42

QY 62 VIVFENHREKEIAVRVLELMGAKRYVYHIIPAADLKVRLVLLVIGSLTGKAKLSGVR 121
Db 43 QSVESY-----DVDVIHEFEIPIVHAELTEELKELNDPNVKA-----83

QY 122 FIQEDYKVTVAE-GLDESAQAQVNAATYVNLGVGSGDITIGIDGIDASHPDLQGV 180
Db 177 QGKVLGWDFYNGRSYDD-HGCTHVASIAAGTGAASNGK--YKGMAPGAKLAGIKVL 233
Db 149 --NVRGASFPVSTNPQOGSGHGV-----AGTIAALNNSIGVLOTYSPASLYAVKVL 202
QY 234 GADGSGSTIIKGVAVDNKDKYIKVINLSGSSQSDGTDLSQAVNNAWDAGIYV 293
Db 203 DSTGSGOYSWIINGIENAIN---NNDVINNSLG---GSPSGTALKTVYDKAVSSGTV 255
QY 294 CVAAGNSCP--NTVYSGPAAASKVITVGAVDSDNDNTASPSRGTADGRKLPKPEVAPGV 351
Db 256 AAGAGNCGSSGSSSTVGPAPKPSITAVGAVNNSNORASFSAGSELD-----VMAPCV 309
QY 352 DIIAPRASGTSNGTPINDYTKASGTSMATPHVSGVALLILOAHPSWTPDKVKTALITETA 411
Db 310 SIQSTLPQGT-----YGAVNGISMATPHVAGVAAALDKKHPITWNAQVDRLESTA 360
QY 412 DIVAPKEIADYAGGRVNYKA 434
Db 361 TVLG-----NSFTYKGLINVQAA 379

RESULT 15
B83756
subtilisin-type alkaline proteinase (EC 3.4.21.-) BH0855 precursor [similarity] - Bac
C/Species: Bacillus halodurans
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C/Accession: B83756
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A/Reference number: A83650; MUID:20512582; PMID:11058132
A/Accession: B83756
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-361 <STO>
A/Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04574.1; GSPDB:G
A/Experimental source: strain C-125
C/Genetics:
A/Gene: BH0855
C/Superfamily: subtilisin; subtilisin homology
C/Keywords: hydrolase; serine proteinase
F:1-25/Domain: signal sequence #status predicted <SIG>

Query Match 14.8%; Score 508; DB 2; Length 361;
Best Local Similarity 33.1%; Pred. No. 9.3e-21;
Matches 144; Conservative 80; Mismatches 131; Indels 80; Gaps 15;

QY 2 KGLKALILVLGLVGSVAAPEKKEVQVRNVEKNYGLLTPGLFRKIQKLPNEEIST 61
Db 3 QSLKVMVLSVAL-LFMANPAAASEEKEYLIVPEP-----DEVA 42

QY 62 VIVFENHREKEIAVRVLELMGAKRYVYHIIPAADLKVRLVLLVIGSLTGKAKLSGVR 121
Db 43 QSVESY-----DVDVIHEFEIPIVHAELTEELKELK-----KDPNVK 82

QY 122 FIQEDYKVTVAE-GLDESAQAQVNAATYVNLGVGSGDITIGIDGIDASHPDLQGV 181
Db 83 AIEKNAEVTIS---QTPWGISFINTQQAHHNG-IFGNGARVAVLDGTI-ASHPDL--RIA 136

QY 182 GWDFYNGRSYDD-HGCTHGHVTHASIAAGTGAASNGK--YKGMAPGAKLAGIKVLGADSG 239
Db 137 GGASFTISSEPSYHDNNGHGVTHV---AGTIAALNNSIGVLOTYSPASLYAVKVLDRNGSG 192

QY 240 SISTIILKGVAVDNKDKYIKVINLSGSSQSDGTDLSQAVNNAWDAGIYVCAAGN 299
Db 193 SLASVAQGIEMAIN---NHHIINNSLG---STSGSSTLEAVNRRANNAGILLVGAAGN 245

QY 300 SGPTNTYVGSPPAAASKVITVGAVDSDNDNTASPSRGTADGRKLPKPEVAVGVDIIAPRAS 359
Db 246 TG--ROGVNYPARYSGVNAVAADQGRASFTYGP-----EIEISAPGVNV-----291

Db 46 TMSANSSAKKD-----VISEKGGKQKQFYVNAATIDKAVREL-----KKD 91
QY 118 SGVRFIOEDYKVTVAE-LEGLDESAQAQVNAATYVNLGVGSGDITIGIDGIDASHPDL 176
Db 92 PSVAYVEDH---IAHEYAQSPYGISQIKAPALHSQGYTGSNVKVVAVJDSGIDSSHPDL 148
QY 177 QGKVLGWDFYNGRSYDD-HGCTHVASIAAGTGAASNGK--YKGMAPGAKLAGIKVL 233
Db 149 --NVRGASFPVSTNPQOGSGHGV-----AGTIAALNNSIGVLOTYSPASLYAVKVL 202
QY 234 GADGSGSTIIKGVAVDNKDKYIKVINLSGSSQSDGTDLSQAVNNAWDAGIYV 293
Db 203 DSTGSGOYSWIINGIENAIN---NNDVINNSLG---GSPSGTALKTVYDKAVSSGTV 255
QY 294 CVAAGNSCP--NTVYSGPAAASKVITVGAVDSDNDNTASPSRGTADGRKLPKPEVAPGV 351
Db 256 AAGAGNCGSSGSSSTVGPAPKPSITAVGAVNNSNORASFSAGSELD-----VMAPCV 309
QY 352 DIIAPRASGTSNGTPINDYTKASGTSMATPHVSGVALLILOAHPSWTPDKVKTALITETA 411
Db 310 SIQSTLPQGT-----YGAVNGISMATPHVAGVAAALDKKHPITWNAQVDRLESTA 360
QY 412 DIVAPKEIADYAGGRVNYKA 434
Db 361 TVLG-----NSFTYKGLINVQAA 379

RESULT 14
A48373
high-alkaline serine proteinase (EC 3.4.21.-) precursor - Bacillus sp. (strain AH-101)
N:Alternate names: subtilisin-like thermostable alkaline serine proteinase
C:Species: Bacillus sp.
C/Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 28-May-1999
C/Accession: A48373; J50714
R:Takami, H.; Kobayashi, T.; Aono, R.; Horikoshi, K.
Appl. Microbiol. Biotechnol. 38, 101-108, 1992
A/Title: Molecular cloning, nucleotide sequence and expression of the structural gene for subtilisin-like thermostable alkaline serine proteinase from Bacillus sp. strain AH-101
A/Reference number: A48373; MUID:93098926
A/Accession: A48373
A/Molecule type: DNA
A/Residues: 1-361 <TA2>
A/Cross-references: GB:S50880; NID:g261737; PIDN:AAC60421.1; PID:g261738
A/Experimental source: AH-101
A/Note: this sequence is inconsistent with the nucleotide translation
R:Takami, H.; Kobayashi, T.; Aono, R.; Horikoshi, K.
submitted to JIPID, July 1992
A/Description: Molecular cloning, nucleotide sequence and expression of the structural gene for subtilisin-like thermostable alkaline serine proteinase from Bacillus sp. strain AH-101
A/Reference number: J50714
A/Accession: J50714
A/Molecule type: DNA
A/Residues: 34-334, L', 336-361 <TA2>
C/Comment: This alkaliphilic Bacillus homolog to the subtilisins of neutrophilic Bacillus
C/Superfamily: subtilisin; subtilisin homology
C/Keywords: extracellular protein; hydrolase; serine proteinase
F:115-321/Domain: subtilisin homology <SBT>
F:124,154,307/Active site: Asp, His, Ser #status predicted

Query Match 14.8%; Score 508; DB 2; Length 361;
Best Local Similarity 33.7%; Pred. No. 9.3e-21;
Matches 147; Conservative 79; Mismatches 128; Indels 82; Gaps 17;

QY 2 KGLKALILVLGLVGSVAAPEKKEVQVRNVEKNYGLLTPGLFRKIQKLPNEEIST 61
Db 3 QSLKVMVLSVAL-LFMANPAAASEEKEYLIVPEP-----DEVA 42

QY 62 VIVFENHREKEIAVRVLELMGAKRYVYHIIPAADLKVRLVLLVIGSLTGKAKLSGVR 121
Db 43 QSVESY-----DVDVIHEFEIPIVHAELTEELKELNDPNVKA-----83

QY 122 FIQEDYKVTVAE-GLDESAQAQVNAATYVNLGVGSGDITIGIDGIDASHPDLQGV 180
Db 177 QGKVLGWDFYNGRSYDD-HGCTHVASIAAGTGAASNGK--YKGMAPGAKLAGIKVL 233
Db 149 --NVRGASFPVSTNPQOGSGHGV-----AGTIAALNNSIGVLOTYSPASLYAVKVL 202
QY 234 GADGSGSTIIKGVAVDNKDKYIKVINLSGSSQSDGTDLSQAVNNAWDAGIYV 293
Db 203 DSTGSGOYSWIINGIENAIN---NNDVINNSLG---GSPSGTALKTVYDKAVSSGTV 255
QY 294 CVAAGNSCP--NTVYSGPAAASKVITVGAVDSDNDNTASPSRGTADGRKLPKPEVAPGV 351
Db 256 AAGAGNCGSSGSSSTVGPAPKPSITAVGAVNNSNORASFSAGSELD-----VMAPCV 309
QY 352 DIIAPRASGTSNGTPINDYTKASGTSMATPHVSGVALLILOAHPSWTPDKVKTALITETA 411
Db 310 SIQSTLPQGT-----YGAVNGISMATPHVAGVAAALDKKHPITWNAQVDRLESTA 360
QY 412 DIVAPKEIADYAGGRVNYKA 434
Db 361 TVLG-----NSFTYKGLINVQAA 379

RESULT 14
A48373
high-alkaline serine proteinase (EC 3.4.21.-) precursor - Bacillus sp. (strain AH-101)
N:Alternate names: subtilisin-like thermostable alkaline serine proteinase
C:Species: Bacillus sp.
C/Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 28-May-1999
C/Accession: A48373; J50714
R:Takami, H.; Kobayashi, T.; Aono, R.; Horikoshi, K.
Appl. Microbiol. Biotechnol. 38, 101-108, 1992
A/Title: Molecular cloning, nucleotide sequence and expression of the structural gene for subtilisin-like thermostable alkaline serine proteinase from Bacillus sp. strain AH-101
A/Reference number: A48373; MUID:93098926
A/Accession: A48373
A/Molecule type: DNA
A/Residues: 1-361 <TA2>
A/Cross-references: GB:S50880; NID:g261737; PIDN:AAC60421.1; PID:g261738
A/Experimental source: AH-101
A/Note: this sequence is inconsistent with the nucleotide translation
R:Takami, H.; Kobayashi, T.; Aono, R.; Horikoshi, K.
submitted to JIPID, July 1992
A/Description: Molecular cloning, nucleotide sequence and expression of the structural gene for subtilisin-like thermostable alkaline serine proteinase from Bacillus sp. strain AH-101
A/Reference number: J50714
A/Accession: J50714
A/Molecule type: DNA
A/Residues: 34-334, L', 336-361 <TA2>
C/Comment: This alkaliphilic Bacillus homolog to the subtilisins of neutrophilic Bacillus
C/Superfamily: subtilisin; subtilisin homology
C/Keywords: extracellular protein; hydrolase; serine proteinase
F:115-321/Domain: subtilisin homology <SBT>
F:124,154,307/Active site: Asp, His, Ser #status predicted

Query Match 14.8%; Score 508; DB 2; Length 361;
Best Local Similarity 33.7%; Pred. No. 9.3e-21;
Matches 147; Conservative 79; Mismatches 128; Indels 82; Gaps 17;

QY 2 KGLKALILVLGLVGSVAAPEKKEVQVRNVEKNYGLLTPGLFRKIQKLPNEEIST 61
Db 3 QSLKVMVLSVAL-LFMANPAAASEEKEYLIVPEP-----DEVA 42

QY 62 VIVFENHREKEIAVRVLELMGAKRYVYHIIPAADLKVRLVLLVIGSLTGKAKLSGVR 121
Db 43 QSVESY-----DVDVIHEFEIPIVHAELTEELKELNDPNVKA-----83

QY 122 FIQEDYKVTVAE-GLDESAQAQVNAATYVNLGVGSGDITIGIDGIDASHPDLQGV 180
Db 177 QGKVLGWDFYNGRSYDD-HGCTHVASIAAGTGAASNGK--YKGMAPGAKLAGIKVL 233
Db 149 --NVRGASFPVSTNPQOGSGHGV-----AGTIAALNNSIGVLOTYSPASLYAVKVL 202
QY 234 GADGSGSTIIKGVAVDNKDKYIKVINLSGSSQSDGTDLSQAVNNAWDAGIYV 293
Db 203 DSTGSGOYSWIINGIENAIN---NNDVINNSLG---GSPSGTALKTVYDKAVSSGTV 255
QY 294 CVAAGNSCP--NTVYSGPAAASKVITVGAVDSDNDNTASPSRGTADGRKLPKPEVAPGV 351
Db 256 AAGAGNCGSSGSSSTVGPAPKPSITAVGAVNNSNORASFSAGSELD-----VMAPCV 309
QY 352 DIIAPRASGTSNGTPINDYTKASGTSMATPHVSGVALLILOAHPSWTPDKVKTALITETA 411
Db 310 SIQSTLPQGT-----YGAVNGISMATPHVAGVAAALDKKHPITWNAQVDRLESTA 360
QY 412 DIVAPKEIADYAGGRVNYKA 434
Db 361 TVLG-----NSFTYKGLINVQAA 379

us-09-841-553-5.rpr

```

QY 360 GTSNGTPINDYTKKSGTSMATPHVSGVGALILQARHSPDVKVTALETADIVAPKSI 419
Db 292 ---NSTYTGRRVVSUSGTSMTAPHVAGVAALVKRSPSYNNQIRORINQATYLGSPSL 348
QY 420 ADIAYGAGRVNYKA 434
Db 349 ----YGNGLVHAGRA 359

```

Search completed: October 31, 2002, 13:31:59
Job time : 21 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 2, 2002, 05:31:18 ; Search time 3395.5 seconds
(without alignments)
16944.853 Million cell updates/sec

Title: US-09-841-553-6
Perfect score: 1977
Sequence: 1 ATGAAGGGGCTGAAAGCTCT.....AGGCGGCTGCTACTACGGG 1977

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054540 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank:
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_cm.*
21: em_or.*
22: em_ov.*
23: em_pet.*
24: em_ph.*
25: em_pi.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pin.*
35: em_htg_rod.*
36: em_htg_man.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1368	58.2	12452	1	AE010265 Pyrococcus
2	1366.4	58.1	1362	6	AR201155
3	1360.8	68.8	1977	6	AR201152 Sequence
4	837	42.3	1236	6	AR201146
5	669.8	33.9	898	6	AR009704 Sequence
6	427	21.6	564	6	AR009707
7	197	10.0	42527	1	SC51A
C 8	195	9.9	14376	1	AE013049 Streptomy
9	191.2	9.7	3010	8	VCA429230
10	141.2	7.1	158063	9	AP001046
11	141.2	7.1	340000	9	AP001751
12	140.6	7.1	87076	9	AC085918
C 13	140	7.1	40899	1	SCC44
C 14	134.8	6.8	169585	2	AC078821
15	131.2	6.6	3348	1	D83672
16	131	6.6	2539	6	AR202322
17	131	6.6	2809	6	AR202321
C 18	128.4	6.5	32704	1	SC8A11
19	126.2	6.4	39369	2	AC115681
C 20	124.4	6.3	97683	2	AC116548
C 21	123.8	6.3	87201	9	AC007623
C 22	123.2	6.2	151705	2	AC117900
C 23	120.6	6.1	176351	2	AC117007
C 24	118	6.0	169163	2	AC115666
C 25	117.8	5.9	181179	9	AL161725
C 26	117.2	5.9	18234	9	CNS05TDS
C 27	116.8	5.9	165337	2	AC096212
C 28	115	5.8	88203	5	AC097628
C 29	112.8	5.7	98734	9	AC013277
C 30	111.4	5.6	115833	2	AC124319
C 31	111	5.6	126599	2	AP003816
C 32	110	5.6	171260	9	AC026743
C 33	109.8	5.6	163034	2	AC099432
C 34	109.6	5.5	150754	9	AC023491
C 35	109.2	5.5	157922	2	AC123764
C 36	109	5.5	167390	9	AC007453
C 37	108.8	5.5	23542	2	AC056320
C 38	108	5.5	292550	1	AP001513
39	108	5.4	5163	6	AR097042
40	107.6	5.4	5318	6	AR097041
C 41	107.5	5.4	58930	2	AC098321
C 42	107.5	5.4	200350	2	AC095469
C 43	107.5	5.4	184929	9	AC058790
C 44	106.4	5.4	1422	6	AX451732
45	106	5.4			

ALIGNMENTS

RESULT 1
AE010265/c
LOCUS Pyrococcus furiosus DSM 3638, section 140 of 173 of the complete genome.
DEFINITION AE010265.1 GI:18893813
ACCESSION AE010265 AE009950
VERSION
KEYWORDS
SOURCE Pyrococcus furiosus DSM 3638.
ORGANISM Pyrococcus furiosus DSM 3638.
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
REFERENCE
1 (bases 1 to 12452)
Maeder, D.L., Weiss, R.B., Dunn, D.M., Cherry, J.L., Gonzalez, J.M.,
DiRuggiero, J. and Robb, F.T.

Thu Nov 7 10:12:15 2002

Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and
P. horikoshii inferred from complete genomic sequences
Genetics 152 (4), 1299-1305 (1999)
10430560
REFERENCE
PUBMED
JOURNAL
MEDLINE
AUTHORS
TITLE
The complete sequence of the *Pyrococcus furiosus* genome
Unpublished
REFERENCE
JOURNAL
MEDLINE
AUTHORS
TITLE
Direct Submission
Submitted (12-FEB-2002) Human Genetics, University of Utah, 20
South 2030 East, Salt Lake City, UT 84112, USA
FEATURES
source
1. 12452
/organism="Pyrococcus furiosus DSM 3638"
/strain="DSM 3638"
/db_xref="taxon:186497"
132. 1013
/gene="PF1657"
132. 1013
/gene="PF1657"
/feature="Function Code: 13.1 Translation: Amino Acyl LRNA
Synthetases; (hisp)"
/codon_start=1
/transl_table=11
/product="histidyl-tRNA synthetase"
/protein_id="AAL81781.1"
/db_xref="GI:18893814"
/translation="MKRSMEEVYKIAETANKLRISFQMGYKVVFFIIEYSPET
RGTNAYNLFILSPDWTSLIKQKDNKAVYIAEVLNGVKGWAGIEFEGG
KGLGYLEALSIATLSLESLGITEYIDIGSIFVWKLSEGVWPKIEALIKRNEFI
ENIPKDKKDWELFNRRGRSIEKLOTIVDSIEDPRVLDLGVRLPYDII
FIISPOVGKPIGGGEYVRGKQIGFVFLNKLUSLYRPKDRARKVISEIDPKAY
STAREMVLGIPVEVRL"
1010. 1624
/gene="PF1658"
1010. 1624
/gene="PF1658"
/feature="Function Code: 1.7 Amino Acid Biosynthesis:
Histidine family; (hisp)"
/codon_start=1
/transl_table=11
/product="Atp phosphoribosyltransferase"
/protein_id="AAL81782.1"
/db_xref="GI:18893815"
/translation="MREVLPRGLYQDSIKVLELAGIKVPKDERALLMGDNKEFL
ARAFDPVVEHGDIGIAGSDVEERGSDVIFLEUPFGKCRSLIAPVPEKVVDEEN
MDGYKIATKPNITKRYFNENIEVEVILKHSIELAPRIGLADALVIVETGNTLES
NGLVEAKIMDVYALLLVNRIISQKMFEEINSFTIKRSVYGH"
1614. 2741
/gene="PF1659"
1614. 2741
/gene="PF1659"
/feature="Function Code: 1.7 Amino Acid Biosynthesis:
Histidine family"
/codon_start=1
/transl_table=11
/product="histidinol dehydrogenase"
/protein_id="AAL81783.1"
/db_xref="GI:18893816"
/translation="WDIEYVEILKDIRERGLAEIREYSKKFDGYEGPFEYSKEEF
EAEERIPERDKEIKRILKRIEYKRLKQDILYENGSLYGLIVKRLKRIYVPG

GRPLSTLMVGVPARVACVKEIVTTPPRDKGVNPVILYIAKLKIEEYKLGGLTOA
IAAAYGIGKHKQKIFGPKNFKNEARQVFGIVGIDSLAGPSIAIVADEADAPFL
VLYDLQLSQRHGDSDKAWLLITTSFKNEADYREGITIVYCSLEECCKANETAPRHL
EITPREPKLVDLLENAGAIYGLIYTPVPAADYFLGNHVLPTGCAKFMGLVLTWDF
HAPISIANVSEEFISERELGIRLAETIEGIELHKKSLVRK"
2808. 3338
/gene="PF1660"
2808. 3338
/gene="PF1660"
/feature="Function Code: 1.7 Amino Acid Biosynthesis:
Histidine family"
/codon_start=1
/transl_table=11
/product="imidazoleglycerol-phosphate dehydratase"
/protein_id="AAL81784.1"
/db_xref="GI:18893817"
/translation="MRRTTKETDITVEIGKKEIKTNDLIDHMLTAFAYLGDMDRI
TATYDRLHMLWEDIGIIGLAEALRENLPEKFPFGNALMPMDALVLSVDSNRPYAN
VDYNIKDAEFGFAVSLLEKFEVWGLANGLEBATIHIXQLSGENAHIVEAFAKGLMALR
VATKESERVESTKGLV"
3338. 3922
/gene="PF1661"
3338. 3922
/gene="PF1661"
/feature="Function Code: 1.7 Amino Acid Biosynthesis:
Histidine family (hisp)"
/codon_start=1
/transl_table=11
/product="glutamine amidotransferase"
/protein_id="AAL81785.1"
/db_xref="GI:18893818"
/translation="MDRIAVDLGIGNLANVKKALKGVITSDPYIEKADKIVLPGVG
NFGAVDKLAPIKDIIIEGINEGPFGLICLQKOLLFEESPEGKGLIGIKRQVVK
LKRVTPHIGMNVKWKLEKLEKNGSYFIFVSHYAVPQDPDIIATITDYNAAE
FVSSVGFENIEGVFHPKSKKGLIILNFRRL"
3924. 4646
/gene="PF1662"
3924. 4646
/gene="PF1662"
/feature="Function Code: 1.7 Amino Acid Biosynthesis:
Histidine family"
/codon_start=1
/transl_table=11
/product="n-(5'-phospho-d-ribose)formimino)-5-amino-1-
(5'-phosphoribosyl)-4-imidazolecarboxamide isomerase"
/protein_id="AAL81786.1"
/db_xref="GI:18893819"
/translation="MKRIYPAIDIMDKVVRVLYKKDKVKKVGDPLKAFKFESEY
DKIHVDLDGAFSGKPKNDIVKIRDVGVVQIGGPGDYTESIKKADIDGVENVII
GPKALDISFLERITSEFGITVSLDSKGGKIFTKWIIEESLLVEDAYKMLRFAVRK
VYTVIDRGTLGLGIEKIKRFGWDEEFIVAGGVATPTDIIQLAIGFSGVIGKALYEG
TISLEALEVAKSVGKNYSSS"
4570. 5370
/gene="PF1663"
4570. 5370
/gene="PF1663"
/feature="Function Code: 1.7 Amino Acid Biosynthesis:
Histidine family; (hisp)"
/codon_start=1
/transl_table=11
/product="imidazoleglycerol-phosphate synthase, cyclase
subunit"
/protein_id="AAL81787.1"
/db_xref="GI:18893820"
/translation="MKSLSLKLLKRLKLVARRIATLIDKIDRVVKKIKFKNIRDA
GDPIELAKRSEIDSIYFLDITASYEKRTLLDLVRAVEIYVFPVGGGIVTVE
EAREIIRGADKIFINAAVENSILKEADKIGLANVVAIDAKWNGFWEVYTHGG
RKPRGDVAVENAKRVEELGAGEILLTSDMDIDGTGKGDDPLDKAVDVANIPVJASGG
AKGPHFLEVPFRIGSDAALAAISIFHYGETTIVRELKWLAKNGIPVRLD"
5345. 5010
/gene="PF1664"
5345. 5010

Thu Nov 7 10:12:15 2002

0:

Matches 984; Conservative 0; Mismatches 245; Indels 0; Gaps 0:

Db	1244	CCCCAAGGAGATAGCGACATCCCTACGGTGGGGTAGGGTCAAGCTCTACAGGCCA	1303
Qy	1304	TCAAGTACGAGACTACGCCAAGCTACCTTCAACGGGTCCGTCCCGACAGGGAAGCG	1363
Db	1304	TCAAGTACGAGACTACGCCAAGCTACCTTCAACGGGTCCGTCCCGACAGGGAAGCG	1363
Qy	1364	CCAGCCACCTTCGAGCTCAGCGGCGCACCTTCGTGACCGCCACCTCTACTGGACA	1423
Db	1364	CCAGCCACCTTCGAGCTCAGCGGCGCACCTTCGTGACCGCCACCTCTACTGGACA	1423
Qy	1424	CGGGTCGAGGACATCGACTCTACCTTACGACCCCAAGGAAGAGCTGACTACT	1483
Db	1424	CGGGTCGAGGACATCGACTCTACCTTACGACCCCAAGGAAGAGCTGACTACT	1483
Qy	1484	CCTACACCGCTACTACGGCTTCGAGAGGCTCGGCTACTACACCCGACCGCGAAGCT	1543
Db	1484	CCTACACCGCTACTACGGCTTCGAGAGGCTCGGCTACTACACCCGACCGCGAAGCT	1543
Qy	1544	GGACGTCAGGTCGTCAGCTACAAAGGGCGGCGGAAGTACCAAGTGCAGCTCGCAGCG	1603
Db	1544	GGACGTCAGGTCGTCAGCTACAAAGGGCGGCGGAAGTACCAAGTGCAGCTCGCAGCG	1603
Qy	1604	ACGGAGGCTCAGCTCAGCTCGGGGCGCAACCGGATCCAAACCCCAACCGCAACAA	1663
Db	1604	ACGGAGGCTCAGCTCAGCTCGGGGCGCAACCGGATCCAAACCCCAACCGCAACAA	1663
Qy	1664	CCCGACACGACACGACCTTCACGGTTCGGTTCAGGACTGAGGACACAGCG	1723
Db	1664	CCCGACACGACACGACCTTCACGGTTCGGTTCAGGACTGAGGACACAGCG	1723
Qy	1724	ACACCTTACCATGAGCTCAACAGCGGTGCGCAACAGATACCGGTGACCTGCTCG	1783
Db	1724	ACACCTTACCATGAGCTCAACAGCGGTGCGCAACAGATACCGGTGACCTGCTCG	1783
Qy	1784	ATATCTTCTACACGCTCGACTCTACCTTACGACCCCAAGCAAGCTGCTGACCA	1843
Db	1784	ATATCTTCTACACGCTCGACTCTACCTTACGACCCCAAGCAAGCTGCTGACCA	1843
Qy	1844	GGTCCAGGTGCGACAGCTACGACGCTGAGTACGACGACGACGACGACGACGAC	1903
Db	1844	GGTCCAGGTGCGACAGCTACGACGCTGAGTACGACGACGACGACGACGACGAC	1903
Qy	1904	GGAGTTCCTCTGCTACGCTACGACGACGACGACGACGACGACGACGACGAC	1963
Db	1904	GGAGTTCCTCTGCTACGCTACGACGACGACGACGACGACGACGACGACGAC	1963
Qy	1964	TCGCTACTACGGG	1977
Db	1964	TCGCTACTACGGG	1977
RESULT 4			
LOCUS	AR201146	1236 bp	DNA
DEFINITION	Sequence 2 from patent US 6358726.	linear	PAT 20-APR-2002
ACCESSION	AR201146		
VERSION	AR201146.1	GI:20252034	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 1236)		
AUTHORS	Takekura H., Morishita M., Shimojo T., Asada K. and Kato I.		
TITLE	Thermolabile protease		
JOURNAL	Patent: US 6358726-A 2 19-MAR-2002;		
FEATURES	Location/Qualifiers		
source	1..1236		
BASE COUNT	368 a 251 c 309 g 308 t		
ORIGIN			
Query Match	42.3%	Score 837;	DB 6; Length 1236;
Best Local Similarity	80.1%	Pred. No. 4e-105;	

```
QY 1477 CACTACCTCCTACACCGCCTACTACGGCTTCGAGAAAGTTCGGCTACTACAAACCGGACGCC 1536
DB 1081 CACTACTCTTACACCGCCTACTACTGATTGAAAGAGTTGTTATTACAAACCAACTGAT 1140
QY 1537 GGAACCTGGACGGTCAAGGTCGTGAGCTACAAGGCGCGGCGCACTACAGGTGCGAGTC 1596
DB 1141 GGAACATGACAAATTAAGTTGTAAGCTACAGCGAAGTGAACATATCAAGTAGATGTG 1200
QY 1597 CTCAGCGCGCGGAGCCTCAGCCAGTCGG 1625
DB 1201 GTAAGTAGTGTTCCTTTCACAGCCTGG 1229

RESULT 5
LOCUS AR009704
DEFINITION Sequence 7 from patent US 5756339.
ACCESSION AR009704
VERSION AR009704.1 GI:3968509
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  Unclassified
  1 (bases 1 to 898)
AUTHORS Mitsu.M., Yamamoto,K., Morishita,M., Asada,K., Tsunasawa,S. and
  Kato,I.
TITLE Hyperthermostable protease gene
JOURNAL Patent: US 5756339-A 7 26-MAY-1998;
FEATURES
  Location/Qualifiers
    source
      1..898
      /organism="unknown"
BASE COUNT 192 a 299 c 266 g 130 t 11 others
Query Match 33.9%; Score 669.8; DB 6; Length 898;
Best Local Similarity 86.3%; Pred. No. 2.8e-82;
Matches 780; Conservative 0; Mismatches 118; Indels 6; Gaps 4;

QY 523 GATCTCCAGGAAGTAATGGGTGGTAGATTTTGTCAATGTAGGAGTATCCATAC 582
DB 1 GATCTGAAGGCAAGGTATAGGCTGTACGACGCGCTCAACGCGAGTGCACCCCTAC 60
QY 583 GATGACATGGACATGGAACCTCATGTCTCAATAGCAGCTGGTACTGGACAGCAAGT 642
DB 61 GATGACAGGACAGGAACCCAGCTTGGGGTATCGTCCGGAACCGGACGCG 117
QY 643 AATGGCAAGTACAAGGGAATGGCTCCAGCAGCTAAGCTGGCGGGAATTAAGGTTCTAGT 702
DB 118 AACTCCAGTACATAGCGCTCGCCCGCGCGAAGCTCTCGCGCTCAAGGTTCTCGT 177
QY 703 CGCGATGTTCTGGAGCATATCTATTAATTAAGGAGTTCAGTGGCGCTTGATAAC 762
DB 178 CGCGAGGTTGGGAAGCGCTCCACCATCATCGCGGTGTTGACTGGGTGCTCCAGAC 237
QY 763 AAGATAGTACGGAATTAAGGTCAATTTCTTCTTGGTCAAGCCAGCTCCGAC 822
DB 238 AAGGACAAGTACGGGATGAAGGTTCATCAACCTCTCCCTCGGCTCCCGAGCTCCGAC 297
QY 823 GGAACCGACTCCCTCAGTCAGCGCTCAACAGCCTGGGACCGCGGTATAGTCTGC 882
DB 298 GGAACCGACTCCCTCAGTCAGCGCTCAACAGCCTGGGACCGCGGTATAGTCTGC 357
QY 883 GTCCGCGCGCGCAACAGCGCGCGCAACCTACACGCTCGGCTACCGCGCGCGGACG 942
DB 358 GTCCGCGCGCGCAACAGCGCGCGCAACCTACACGCTCGGCTACCGCGCGCGGACG 417
QY 943 AAGGTCAATACCGTCTGGTGTGACAGCAACAGCAATCGGACGCTTCCAGCAGG 1002
DB 418 AAGGTCAATACCGTCTGGTGTGACAGCAACAGCAATCGGACGCTTCCAGCAGG 477
QY 1003 GCACCGACCGGACGAGGCTCAAGCGGGAAGTGTGCGCCCGCGCTTGACATCATA 1062
  |||||||
```

```
DB 478 GGACCGACCGCGCACGGAAGGCTCAAGCGGAAAGTCTGTCGCCGCCCGCGCTTGACATCATA 537
QY 1053 CCCCGCGCGCGCACGGAAGCAGCAGTGGGACACCGCGATAAACGACTACTACACCAAGGCC 1122
DB 538 GCCCGCGCGCGCACGGAAGCAGCAGTGGGACACCGCGATAAACGACTACTACCAAGGCC 597
QY 1123 TCTGGAAACAGCATGGCCACCGCGCACGCTTTGGGGCGTTGGCGGCTCATCTCCAGGCC 1182
DB 598 TCTGGATCCAGCATGGACACCGCGCACGCTTTGGGGCGTTGGCGGCTCATCTCCAGGCC 557
QY 1183 CACCGAGCTGGACCGCGGACAAAGGTGAAGACCGCGCTCATCGAGCCCGCGCATAGTC 1242
DB 658 CACCGAGCTGGACCGCGGACAAAGGTGAAGA-CGCGCTCATCGAGCCCGCGCATAGTC 716
QY 1243 GCCCGCAAGGAGATAGCGGACATCCGCTACGGTGGGTAGGCTGAACGCTCAACAGGCC 1302
DB 717 GNCGCCAAGGAGATAGCGGACATCCGCTACGGTGGGTAGGCTGAACGCTCAACAGGCC 776
QY 1303 ATCAAGTAGGACGACTACGCCAAGCTCACCTTCACCGGCTCCGTCGCGGACAAAGGAGG 1362
DB 777 ATCAAGTNCAGGACTACGNCAGNTCACTTCACCGGNTCCGTCGCGGACAAAGGAGG 836
QY 1363 GCCACCCACACCTTCGACGCTCAGGGCGCGCACCTTCGTCACCGCCACCCCTCTACTCGGAC 1422
DB 837 GGCA-CCACACCTTCGACGCTCAGNGCGGGA-CTTCGTAACGACGACCCCTCTCTNGGAC 894
QY 1423 ACGG 1426
DB 895 ANGG 898

RESULT 6
LOCUS AR009707
DEFINITION Sequence 11 from patent US 5756339.
ACCESSION AR009707
VERSION AR009707.1 GI:3968512
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  Unclassified
  1 (bases 1 to 564)
AUTHORS Mitsu.M., Yamamoto,K., Morishita,M., Asada,K., Tsunasawa,S. and
  Kato,I.
TITLE Hyperthermostable protease gene
JOURNAL Patent: US 5756339-A 11 26-MAY-1998;
FEATURES
  Location/Qualifiers
    source
      1..564
      /organism="unknown"
BASE COUNT 121 a 195 c 161 g 87 t
Query Match 21.5%; Score 427; DB 6; Length 564;
Best Local Similarity 86.1%; Pred. No. 4.3e-49;
Matches 485; Conservative 0; Mismatches 75; Indels 3; Gaps 1;

QY 595 CATGGAAGTCTAGTCTCAATAGCAGCTGGTACTGGACGCAAGTAATGCAAGTAC 654
DB 4 CAGGGAAGTCACTGGCGGAACAGTTGCCGGAACAGGACGCG---TTAACTCCAGTAC 60
QY 655 AAGGGAATGCGCTCCAGAGCTAAGCTGGCGGAATTAAGGTTCTAGTGGCGGATGTTCT 714
DB 61 ATAGGCGTCCGCCCGCGGCAAGCTCGTGGTCAAGGTTCTCGGTCGCGGAGGTTGG 120
QY 715 GGAAGCATATCTACTATTAATTAAGGAGTTAGTGGCGGCTTGATATAAAGATAAGTAC 774
DB 121 GGAAGCTCTCCACCATCATCGCGGTTGTTGACTGGTGGTCCGAGAACAGGATAAGTAC 180
QY 775 GGAATTAAGTCAATTAATCTTCTTCTTGGTTCAAGCCAGAGCTCCGACGGAAGGACTCC 834
DB 181 GGAATTAAGGCTCAATCAACTCTCCCTCGGCTCTCCAGAGCTCCGACGAGGAGGACTCC 240
QY 835 CTCAGTCAGCGGTCAACACGCGCTGGGACGCGGCTATAGTCTGCGTCCGCCCGCGC 894
```

Thu Nov 7 10:12:15 2002

in codons is given for each CDS.
usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov.jp/>
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, ggg, tgg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. Cosmid 51A overlaps cosmid 776 on the AseI-F genomic restriction fragment.

FEATURES
SOURCE
1. 42527
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 51A"
complement(13..2310)
/gene="SCF51A.01c"
/note="SCO0423"
complement(13..2310)
/gene="SCF51A.01c"
/note="SCF51A.01c"
len: 765 aa. Similar to many Streptomyces coelicolor integral membrane proteins including TR:O53902 (EMBL: M64683) ORF1.4 (ACT11) (711 aa), fasta scores opt: 251 z-score: 295.3 E(1): 4.3e-09 27.7% identity in 779 aa overlap, TR:Q92577 (EMBL: AL035569) SC8D9.14 (748 aa), fasta scores opt: 1021 z-score: 1028.4 E(1): 0.40.6% identity in 773 aa overlap and TR:O88022 (EMBL: AL031107) SC5A7.16C (705 aa), fasta scores opt: 639 z-score: 845.0 E(1): 1.4e-28 37.2% identity in 764 aa overlap. Contains multiple possible transmembrane hydrophobic domains."

trans_start=1
/codon_start=1
/product="putative membrane transport protein"

/product="putative membrane transport protein"
/protein_id="CAB56653.1"
/db_xref="GI:6002209"
/db_xref="SPTREMBL:Q9RL83"
/translation="NATLLRIGRGAFFRRRWYTLAAVYLVGVVILKMSAGAGCDDSDMPQIESQAFDLINERPKAEAGADARIVLVPDQGVQVTAETRAADITLVAVADGPOVAASDPLDAATVSDQSTAYATVETEDNVDAGADITLAVRATKISIVVAGGAVAGEPPAGIGEPGLGVAAVLLITFGSLAAGLPLITAVVGLTASIVALGSLGNTASGDLAMVIGIAGVGVYVLLVSVRYREARHDAAGLAVGTAGSAYRAGLVVYALAGLSVGVPSLTKMGLAAAVVAVLITLVLPALCGPFPDALLARYRKRGRSAPARRKRGKSTPPARRLGRGPGKGAAREAGANSVVRARVORRPVVLGADDPKSAVAASDLAATAGVSVVAPPRNAKDTAVIOAVPATPATERTEELVHIDRIRHATESATGAIFETCTTALNDVARKMTDALIPYLLVVGGLAVLIVFRSVLPLKAAGLFSLVGSFGLVTFIQNGHGGGLGVEETGPMVMTPLMLVGI VFGLANDQVFLVSRMREAVFVRPAEAVYVTFGRSGRVYTAANVINTSVFAGMTGGETLLKMYGVLAVAVLDFAVYVMAFVFAVLVLLGLRANWLPRLRLDLPRLVDVEGEALGRTPAPALAAVPTDRQDTVRA"

2423..2428
2431..2922
/gene="SCF51A.02"
/note="SCO0424"
2431..2922
/gene="SCF51A.02"
/note="SCF51A.02, unknown, len: 168 aa."
/codon_start=1
/trans_start=1
/product="hypothetical protein SCF51A.02."

RBS
gene
CDS

Db 241 CTCAGTCAGGCGTCAACACAGCGCTGGGACGCCGGTATAGTAGTCGTCGCCGCCGCGC 300
QY AACAGCGGCGGCGAACAAC 954
Db 301 AACAGCGGCGGCGAACAAC 360
QY GTCGGTGCAGTTGAC 1014
Db 361 GTCGGTGCAGTTGAC 420
QY GACGGAAGGCTCAACCGCGGAGTCTGCGCCCGCGCGCTGCACATCATAGCCCGCGCGCGC 1074
Db 421 GACGGAAGGCTCAACCGCGGAGTCTGCGCCCGCGCGCTGCACATCATAGCCCGCGCGCGC 480
QY AGCGGAACGAGTGGGACCGCGGATTAACGACTACTACACACAGCGCTTGGACACGACG 1134
Db 481 AGCGGAACGAGTGGGACCGCGGATTAACGACTACTACACACAGCGCTTGGACACGACG 540
QY ATGGCCACCGCGGACGCTTTCGGG 1157
Db 541 ATGGCCACCGCGGACGCTTTCGGG 563

RESULT 7
SC51A
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS

Streptomyces coelicolor cosmid 51A.
AL121596.2 GI:20520772
50S ribosomal protein; aminotransferase; Beta-glucosidase;
binding-protein dependent transport protein; hydrolase; integral
membrane protein; LacI family transcriptional regulator;
MAR-family protein; membrane transport protein;
pyrrolidone-carboxylate peptidase; secreted peptidase; secreted
protein; secreted solute-binding lipoprotein; SIR2-like regulatory
protein; sugar phosphotransferase; tetr family transcriptional
regulator; transcriptional regulator; transmembrane efflux protein.
Streptomyces coelicolor A3(2).
Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 42527)
Kinashi, H. and Hopwood, D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351
8843436
2 (bases 1 to 42527)
Murphy, L. and Harris, D.
Unpublished
3 (bases 1 to 42527)
Thomson, N.R., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
Direct Submission
Submitted (28-SEP-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
On May 9, 2002 this sequence version replaced gi:6002208.

Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/projects/S.coelicolor/>) CDS are
numbered using the following system eg SC7B1.01c, SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length

```
gene
/db_xref="GI:6002210"
/translation="MPONTAGLPLPKRPRGHGAGCDPDGPRVSAVPAACVCEVHGPG
RRRGPRVATATARGPTVRPGPHOICGGLPEHRORLQIAVLGKRAEFEEAPGROVAAA
RATRAPASGRSAGTRIASSVOVPGPARSRPRGAVGNGPRHRAAGRDYALAA
ROA"
complement(3024..3974)
/gene="SCF51A.03c"
/notes="SCF51A.03c"
/transl_table=11
/product="hypoetical protein SCF51A.03c"
/protein_id="CAB56655.1"
/db_xref="GI:6002211"
/db_xref="SPTREMBL:O9RL61"
/translation="MTLSLTHVGHDRVTVAHGAAGTGLFSTVYRPEAAWEAPKPLRP
VRLSGAVYDYPNDHRMHKGLQLTASHLSGONLWGNTYVHGEGYRALPRVSNM
HVSFEETGADHAVIAELRLTWHPHGELWAEERIEVRVDVDTGSGMLTWTNVT
NRREPRFGSPPTTHGRPAAGYTGLEWRGPAPRGGRVFTQASASAESEADAGCPH
GSSARMLAYVGEHDGHDGHTLVFEHAPSNDSHGEGTHPAHWEVRSDPPEAVAFSWA
FHEBLVLPGETLARAYRVVADGAWDRARVAHLAHLAW"
complement(3979..3984)
4159..4161
/transl_table=11
/product="hypoetical protein SCF51A.04"
/protein_id="CAB56656.1"
/db_xref="GI:6002212"
/db_xref="SPTREMBL:O9RL60"
/translation="MEIPLRIELRRAVEADNRVSRNGENATVALAWNSDATAALA
RGIAAYADGTGVPHREVELAVQATGRHIPAR"
4475..4478
4483..5319
/gene="SCF51A.05"
/notes="SCF51A.05"
4483..5319
/gene="SCF51A.05"
/notes="SCF51A.05, possible hydrolase, len: 278 aa. Similar
to many e.g. Pseudomonas fluorescens SK_PRYC_PSEFL
(EMBL:AF031153) non-heme chloroperoxidase (EC 1.11.1.10)
(chloride peroxidase), fasta scores opt: 243 z-score:
282.1 E(); 2.4e-08 28.4% identity in 275 aa overlap and
many putative hydrolases: Streptomyces coelicolor
TR:CA844393 (EMBL:AL078610) putative hydrolase SCH35.34C
(269 aa, fasta scores opt: 259 z-score: 300.1 E());
2.3e-09 29.7% identity in 269 aa overlap. Containing a pfam
match to entry PF00561 abhydrolase, alpha/beta hydrolase
fold"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="CAB56657.1"
/db_xref="GI:6002213"
/db_xref="SPTREMBL:O9RL59"
/translation="MTERLAVDGGTIAIVETVSGPLIVLAHGMDSRAAYRTVVPOLV
AAKRAVADLRGCGESSVDMPKSRATIAQDLIALIRHLGGPVAVYGHVSGSRAAI
AAAREPLVAVYALAPETRKQSVRLGLDLRVKRRFRRLRLIGAGFSGSLPWRSLD
VAYPCVPAHWSRGLRIEISLREPRGMKQANGRSAPTDAAGQLGNVRCPLVYVNG
TLDPDNADPHAEAGVAVADLPISGLGRLEMTGEGHYPHDQEPDEVAALVAFURSDAA
```

```
misc_feature
RA*
4624..5295
/gene="SCF51A.05"
/notes="PFam match to entry PF00561 abhydrolase, alpha/beta
hydrolase fold, score 74.40, E-value 2.4e-18"
5335..5339
5342..5875
RBS
gene
Query Match
Best Local Similarity 10.0% Score 197; DB 1; Length 42527;
Matches 629; Conservative 0; Mismatches 565; Indels 33; Gaps 6;
QY 412 CTGGATAGTCTGACGCTCAAGTTATGACACTTGTGGTGGGAACTATGGAATATGATGCT 471
DB 9882 CTGGACCGCACTAGCGCCAGATCGCGCCGCGAGCTGTGCTCGCGCGCTTACCGGGC 9941
QY 472 TCTGGAATCAATAGGAATATTGACACTGGAATTCAGCTTCTCATCTCCAGATCTCCAA 531
DB 9942 GACGCGCTCAAGGTCGCGCTCTTGACACCGGTGCGGACGAGCCATCCCGACCTGGCC 10001
QY 532 GGAAGATTAATGCTGGTGGTGTAGATTTTGTCAATGCTAGGAGTTATCCATACGATGACCA 591
DB 10002 GCGCGG---TCGCGCGCGCAAGGACTTCTCGCGAGCTCCGCGCAGCAACGACGCTTTC 10058
QY 592 GGACATGGAATCTGTAGCTTCAATAGCACTGTGTACTGGAGCAGCAAGTAATGCAAG 651
DB 10059 GGGCAGCGCAACCCAGCTCGCTCCATCTGTCGGCGGCGCGCGCGCTCCGCGGGAGC 10118
QY 652 TACAAGGAATGGCTCCAGGAGCTTAAGCTGGCGGAGNTTAAGTTCATAGTGCAGAT 711
DB 10119 CGTCAAGGAGTCCGCGCGCGCGCGCTGCTGTCGGCAAGGTCTCGCGGACGACGCG 10178
QY 712 TCTGGAGCATATCTACTATTAATTAAGGAGTTGAGTGGCGCTTGTATACAAAGATAAG 771
DB 10179 TCGGACGCGATCCGAGGTGATCGCGGCGATGAGTGGCGCGG-----AC 10226
QY 772 TACGGAATTAAGTCAATTAATCTTCTTCTTCAAGCCAGAGCTCCGACGAGACGAC 831
DB 10227 CAGGGCGCGAGCTGCTGAACATAGCTCTCGCTCTCGGGGGCGACCGACGACCGAT 10286
QY 832 TCCTCTCAGTCAGCGCGCTCAACAA---CGCTGGGAGCGCGGTATAGTCTGCTGCC 888
DB 10287 CCATCAGCAGGCACTCAACGACCTCACCGCGCACCGCAGCTGTCTGTGTCGCG 10346
QY 889 GCGGCAAC---ACGCGGCGACACCTACACCTGCTCGCTCACCGCGCGCGGAGCAAG 945
DB 10347 GCGGCGACGAGGCGGAGAGGCTCCCGCAGGTCTCGCTCGCGCGCGCGCGCGCG 10406
QY 946 GTCATACCGTCTGTCAGTTACAGCAACGACCAACATCCGCAAGCTTCTCCAGAGGGGA 1005
DB 10407 GCACTCACCGTGGGGCGCTGAGACCGGAGCATTCCTCGCGCGCTTCTCCAGCGCGCG 10466
QY 1006 CGG---ACCGGGGAGCGGAGGCTCAAGCGCGAGCTGCTCGCGCGCGCGGTGACATATA 1062
DB 10467 CGCGGCTCGTGTAGCAGCGCGGTGAAGCGCGACCTCAGCGCGCGCGGTGCGGATCGT 10526
QY 1063 GCGCGCGCGCGCGGACGACGATGGGCAACCGGACCGGATAAAGACTACTACACCAAGGCC 1122
DB 10527 GCGGCGCGCGCGCGCGCGGTGCGCGACCGCTGGCGACCGCTCGACGAGCACTACNCGCGCG 10586
QY 1123 TGTGAACAGCATGCGCACCCCGCACCTTTCGGGGGTTCGGCGCGCTCATCTCCAGGCC 1182
DB 10587 TCGGCGACCTCGATGGCCACCGCGCGCGCGCGCGCGCTCTCTGGCGCGCGCT 10646
QY 1183 CACCGAGCTGGACCGCGCGAGGTGAAGACGCGCTCATCTAGAGCGCGCGAGATGTC 1242
DB 10647 CATCCGACCTGGACCGCGCGCGGTGAAGAGCGCGCTCATCAGCAGCGCGGTGCGGTG 10706
QY 1243 GCGCGCGAGGAGATAGCGGACATCGCTTACGGTGGCGGTGAGGTGAAGC-----TC 1293
DB 10707 GACGCGCGAGGAGTACCGCAACAGCGCGCGCGCGGTGACGTGCGCGCGCGCGACTG 10766
QY 1294 TACAAGGCCATCAAGTAGCAGCACTACGCCAAGCTTCAACCTTCACCGCGCTCGGTGCGCGAC 1353
```

Db 10767 GCGGGTGACCGCCACCGACACTCGTCATGGGCCGCTTCACTCGCGGACACCGAG 10826
QY 1354 AAGGAGCGCCACCCACACCTTGAGGTGAGCGCGCCACCTCGTGACCGCCACCCCTC 1413
Db 10827 CCGGTGAGCGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 10886
QY 1414 TACTGACGACGCGCTCGAGCGACATCGACCTTACTCTAGCACCCACGACGACGAG 1473
Db 10887 GCGGTGAGCGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 10946
QY 1474 GTTACTACTCTACTACCGCTACTACCGCTTCTGAGAAGGTGGCTACTACACCGGAC 1533
Db 10947 TCGACTCGGTACCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 11006
QY 1534 GCGGACCTGAGGTCAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 1593
Db 11007 GCT 11066
QY 1594 GTGTCAGCGACG 1620
Db 11067 GTGCT 11093

RESULT 8
LOCUS AE013049 14376 bp DNA linear BCT 09-MAY-2002
DEFINITION Thermoanaerobacter tengcongensis strain MB4T, section 76 of 244 of the complete genome.
ACCESSION AE013049 AE008691
VERSION AE013049.1 GI:20515814
KEYWORDS
SOURCE Thermoanaerobacter tengcongensis.
ORGANISM Thermoanaerobacter tengcongensis
Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
Thermoanaerobacteraceae; Thermoanaerobacter.
REFERENCE 1 (bases 1 to 14376)
Bao, Q., Tian, Y., Li, M., Xu, Z., Xuan, Z., Hu, S., Dong, W., Yang, J., Chen, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y., Ling, L., Tan, H., Chen, R., Wang, J., Yu, J. and Yang, H.
A complete sequence of the "T. tengcongensis" Genome
Genome Res. 12 (5), 689-700 (2002)
2 (bases 1 to 14376)
Bao, Q., Xu, Z., Hu, S., Dong, W., Chen, Y., Wang, J., Yu, J. and Yang, H.
Direct Submission
Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and Bioinformatics Center, Institute of Genetics and Development, Chinese Academy of Sciences, Beijing Airport Industrial Zone B6, Beijing 101300, China
3 (bases 1 to 14376)
Li, W., Xuan, Z., Yang, J., Ling, L. and Chen, R.
Direct Submission
Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
4 (bases 1 to 14376)
Tian, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y. and Tan, H.
Direct Submission
Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, China
FEATURES
source
1..14376
/organism="Thermoanaerobacter tengcongensis"
/strain="MB4T"
/db_xref="taxon:119072"
265..2374
/gene="GlnA"
/note="TTE0821"
265..269
/gene="GlnA"
278..2374
/gene="GlnA"

/note="Best Blastp hit - gi|2895904|gb|AAC03061.1| (AF046916) glutamine synthetase type III [Ruminococcus flavefaciens], score 851, E-value 0."
/codon_start=1
/transl_table=11
/product="Glutamine synthase"
/protein_id="AAW24078.1"
/db_xref="GI:20515815"
/translation="MQHTLSEIFGNSVRELPKATYKALRKTIIDGIEPLEPA
VAEVANAKMDIAETGAHFQPLGTITAEKRNRSISPTQDGKVIITFSKELI
KEPDASSFGSGRLTTPFARGYTAWDCTSPAFIWDNLITPTACSTGTGALDKIP
LRSMEALSKOALRVLRFGNTAKKVIPTVGPEQYFLIDKMDIKRDLILGRL
FARGSPQGMEDHYFASIKERISAFMDLDELKLEKFPAGINSCKHNSMSTDDQNL
YNTIASDHNLTMELMKVALRHGLVCLLHKEKFPAGINSCKHNSMSTDDQNL
LPGHTPHENQFLVFLCAIKRAIDYADLVRAAATPGNDHRLGANEPATVSIPL
GQDLIDLEQIENGATTSKVGVLKGVSTLPALPKDSTDRNRTPSPFAFGNKPEFR
NGVSSSLATATFLINTVAESIEADLEKADNEVQLLQEIYVKKHRIIFNG
NGISEWKEKEKGLNIBSTVAIPALIKENKVKLMKHGVI.SKRELEARYEVLE
NYIKTINLALTMIDIAKROLPAVNVFATKIAESINSVRATGLNVDISAQTELLAEV
SLMSEFKNISSELENVNEASNMNSDSYKACIYRDVVFVKMGLIREIGDKLETIVD
AELWPLPTVADMLFNI"
842..1279
/gene="GlnA"
/note="Pfam match to entry glin-synt, Glutamine synthetase, score 35.5, E-value 8.90E-11"
complement(2455..2681)
/note="TMR027, Identity:88%, copy 9"
2684..2705
/note="putative rho-independent transcription terminator"
complement(2775..2791)
/note="putative rho-independent transcription terminator"
complement(2778..3360)
/gene="Pdx2"
/note="TTE0822"
complement(2778..3347)
/gene="Pdx2"
/note="Best Blastp hit - gi|17447528|pir||H72371
hypothetical protein TMO472 - Thermotoga maritima (strain MSB) gi|4980982|gb|AAD35557.1|AE001725_2 (AE001725)
amidotransferase, putative [Thermotoga maritima]', score 158, E-value 6.00E-38"
/codon_start=1
/transl_table=11
/product="Predicted glutamine amidotransferase involved in pyridoxine biosynthesis"
/protein_id="AAW24079.1"
/db_xref="GI:20515816"
/translation="MRVGLVQGVSKVHKMDKLUKLIKIEAVEAKDKOTLLTDLALII
PGGSGTAIGKILVDFGLKDAIKLNERKVPWGTGACMLMAKHIVNDRHRLGIMDI
TVVRNAYGSDINSFKRLIIPALISENIEAFIRAPYIESVGVGVRLARVEKRIVAA
QQDLATAPHELDLDDLSFYRVFLRLNS"
complement(2781..3335)
/gene="Pdx2"
/note="Pfam match to entry SNO, SNO glutamine amidotransferase family, score 236.1, E-value 4.90E-67"
complement(3349..4239)
/gene="SnaI"
/note="TTE0823"
complement(3349..4227)
/gene="SnaI"
/note="Best Blastp hit - gi|1176057|sp|P45293|YG47_HAEIN
HYPOPHETICAL PROTEIN HI1647 gi|1074877|pir||F64173
hypothetical protein HI1647 - Haemophilus influenzae (strain Rd KW20) gi|1574495|gb|AAC23294.1| (U32837)
conserved hypothetical protein [Haemophilus influenzae Rd], score 392, E-value 1.00E-108"
/codon_start=1
/transl_table=11
/product="Pyridoxine biosynthesis enzyme"
/protein_id="AAW24080.1"
/db_xref="GI:20515817"
/translation="MNERVELNKNLAKMLKGVIMDVTPPEATIAEKAGAVYNALE
RVPADIRARGGVARNSDPKIIEKKAASIPVMAKVRIGHEVQALGIDFIDIES

[illegible]

Qy	814	ACGTCGACGGACCGAGCTCCCTCAGTCAGGCGCGTCAACAAACGCTGGGAGCGCGGTATA	873
Db	68740	ATCACCAACCATCACCACCATCAGCATCACCACCACTATCACCACCACTATCACCACATC	68799
Qy	874	GTAGTGTGGTGGCGCGGGACAGGGGGGAGACACTTACCGTGGGTGCTCAACCGGC	933
Db	68800	ATCACCAACCATCAGTATCCACACCAACCAACCAACCAACCATCACCACCATCACC	68859
Qy	934	GCGCGAGCAAGGTGTATACGGCTGGTGCAGTTCAGAGCAAGCAACATCGGCACGCTTC	993
Db	68860	ACCACCAACCATCAGCAGCATCAGCATCACCACCATCAGCATCACCACCATCACCATCACC	68919
Qy	994	TCCAGCAGGGGACGACGCGCGGAGCGAGGCTCAAGCGGAGTGGTGGCGCGCGGGT	1053
Db	68920	ACCATCACCATCAGCAGCATCACCACCAACCAACCAACCAACCATCAGCAGCATCACCAC	68979
Qy	1054	GACATATAGCCCGCGCGGACCGGAGCAAGCATGGGACACCGCATAAAGCACTACTAC	1113
Db	68980	ACCACATCAGCAGCATCAGCAGCATCAGCAGCATCAGCAGCATCAGCAGCATCAGCAGCAT	69039
Qy	1114	ACCAGGGCTGTGGACCAAGCATGCGCACCGCGGAGTGTGGGGGTGGGGGCTGATC	1173
Db	69040	ACCACCAACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATC	69099
Qy	1174	CTCAGGCGCCACCGAGCTGGACCGCGGACAGGTGAGACGGCGCTCATCGAGACCGCC	1233
Db	69100	ACCATCAGCATCACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	69159
Qy	1234	GACATAGTGGCGCGCGGAGATAGGCGACATCGCTACGGTGGGGGTAGGCTGAGGTC	1293
Db	69160	ACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	69219
Qy	1294	TACAAGGCCATCAAGTACGAGCGCTAGCGCCAGCTCAGCTTACCGGGCTCGTCCCGAC	1353
Db	69220	ATCAGCAGCATCACCATCACCACCATCACCACCATCACCACCATCACCACCATCACCACCAT	69279
Qy	1354	AAGGAGAGCGCGACCCACACCTTCGAGGTGAGGGGGCGCTTGGTGGACCGCACCTT	1412
Db	69280	ACCACATCAGCAGCACCAACCATCCGACCAACCAACCAACCAACCAACCAACCAACCAAC	69339
Qy	1413	CTACTGGGACAGGGGTGAGGAGCATGAGCACTTACGCTACGACCGCAACGGGAAC	1470
Db	69340	ACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	69399
Qy	1471	GAGTGTGACTCTCTACACCGCTTACTAGGGTTCGAGAGGTGGGTACTACACCGG	1530
Db	69400	ACCACCAACCATCAGCATCACCACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	69459
Qy	1531	ACCGCGGAGCTGGGCGGTGAGTGTGAGCTACAAGGGCGGGGGAACCAACCAACCAACCA	1590
Db	69460	ACCATCAGCAGCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	69519
Qy	1591	GAGTGTGTGAGGAGGGAGCCTCAGCAGTCCGCGGGCGGGAACCAACCAACCAACCAACCA	1650
Db	69520	ACCACCAACCATCACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	69579
Qy	1651	AACCCGAACCCCAACCGGACCCGAGACCCAGACCTTACCGGTTCAGCACTAC	1710

Thu Nov 7 10:12:15 2002

us-09-841-553-6.rge

```

* e.mail: hattori@gs.c.riken.go.jp
* URL: http://hgp.gs.c.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
  Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Molecular Biology, * Tokyo
  160-8582, Japan, r3@db.med.keio.ac.jp
* e.mail: nshimizu@db.med.keio.ac.jp/
* URL: http://www.db.med.keio.ac.jp/
and
* GfK, Dept. of Genome Analysis,
  Mascheroder Weg 1, D-38114 Braunschweig, Germany, * e.mail:
  info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
  Ihnestr. 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
  All63296: Submitted (10-Apr-2000).
  Location/Qualifiers
    1..340000
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="21"
    1..158063
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="21"
      /map="21q22.3"
      /clone_lib="Keio BAC library"
      /note="Accession No. AP001046"
      151173..193275
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="21"
        /map="21q22.3"
        /clone="KB121662"
        /clone_lib="Keio BAC library"
        /note="Accession No. AP001047"
        192809..209397
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /chromosome="21"
          /map="21q22.3"
          /clone="KB121662"
          /clone_lib="Keio BAC library"
          /note="Accession No. AP001048"
          206177..242765
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q22.3"
            /clone="KUD41H11"
            /clone_lib="KUD21-D Cosmid library"
            /note="Accession No. AP001049"
            222082..321579
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="21"
              /map="21q22.3"
              /clone="KB953G5"
              /clone_lib="Keio BAC library"
              /note="Accession No. AP001050"
              293727..>340000
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="21"
                /map="21q22.3"

```

```

/clone="KB161B12.3' partial"
/clone_lib="Keio BAC library"
/note="Accession No. AP001051"
1802..1827
/note="(TTCA)n"
/rpt_family="Simple_repeat"
/rpt_type="TANDEM"
complement(3874..4044)
/note="MERSA"
/rpt_family="DNA/MER1_type"
/rpt_type="DISPERSED"
complement(4382..4654)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type="DISPERSED"
4736..4986
/note="AluX"
/rpt_family="SINE/Alu"
/rpt_type="DISPERSED"
complement(5691..6047)
/note="MLT1A1"
/rpt_family="LTR/MaLR"
/rpt_type="DISPERSED"
complement(11209..11341)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type="DISPERSED"
complement(11335..11465)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type="DISPERSED"
complement(11839..12177)
/note="MLT1A1"
/rpt_family="LTR/MaLR"
/rpt_type="DISPERSED"
complement(12280..12538)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type="DISPERSED"
12539..13330
/note="(GGGA)n"
/rpt_family="Simple_repeat"
/rpt_type="TANDEM"
13365..13608
/note="(GGGA)n"
/rpt_family="Simple_repeat"
/rpt_type="TANDEM"
complement(15402..15494)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type="DISPERSED"
15947..16036
/note="L2"
/rpt_family="LINE/L2"
/rpt_type="DISPERSED"
16175..16347
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type="TANDEM"
16371..16545
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type="TANDEM"
16546..16804
/note="LIMB5"
/rpt_family="LINE/L1"
/rpt_type="DISPERSED"
16805..17026
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type="TANDEM"
17043..17101
/note="(TCCA)n"

```

FEATURES

source

source

source

source

source

source

source

```

/rpt_family='Simple_repeat'
/rpt_type=TANDEM
repeat_region
complement(17104, .17495)

```

Thu Nov 7 10:12:15 2002

Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A.,
 Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M.,
 Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K.,
 Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginder, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulme, W.,
 Iliev, I., Johnson, R., Jones, C., Karatas, A., Lacroque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 MacLean, C., MacDonald, P., Marquis, N., Matthews, C., McCarthy, N.,
 McEvan, P., McKernan, K., McPheeters, R., Meldrum, J., Menes, L.,
 Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
 Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Scanan, S.,
 Severy, P., Sougnez, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, S.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (26-JUN-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 26, 2001 this sequence version replaced gi:13491253.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

 Center: Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu

 Project Information
 Center project name: L394
 Center clone name: 74_R_8

 Only 88.4 kilobases from the middle of the clone are being
 submitted. The remainder overlaps either accession number AC005988
 (WICGR project L484) or accession number AC006947 (WICGR project
 L492).
 FEATURES
 source
 1. .87076
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /clone="RP11-74H3"
 /clone_lib="RPC1-11 Human Male BAC"
 725..793
 /rpt_family="MIR"
 complement(939..1100)
 /rpt_family="AluSg/x"
 complement(1244..1540)
 /rpt_family="AluJb"
 2053..2247
 /rpt_family="MER20"
 3782..3961
 /rpt_family="(CCA)n"
 3980..4900
 /rpt_family="(CCA)n"
 4332
 /note="Probably G, maybe A"
 4358..4561
 /note="Less than 30 qual SNGL region"
 4470..4472
 /note="Fingerprint makes us suspect that there are 100-200
 more ACC repeats missing from here."
 4731..4755
 /note="Less than 30 qual SNGL region"
 4778..4791
 /note="Less than 30 qual SNGL region"
 4802..4808
 /note="Less than 30 qual SNGL region"
 4919..5098
 /rpt_family="(CCA)n"
 8334..8435
 /rpt_family="L1MC4"
 complement(8438..8512)
 /rpt_family="NADE1"
 8576..8861
 /rpt_family="AluSx"
 8878..9010
 /rpt_family="L1MC4"
 complement(9863..10171)
 /rpt_family="AluSg"
 11094..11116
 /rpt_family="AT-rich"
 12327..12483
 /rpt_family="L1PA4"
 13743..13926
 /rpt_family="MIR"
 15390..15928
 /rpt_family="L1ME3A"
 16156..16224
 /rpt_family="L1ME"
 17256..17349
 /rpt_family="MIR"
 17352..17649
 /rpt_family="AluY"
 18933..19226
 /rpt_family="AluSx"
 complement(19402..19574)
 /rpt_family="MER45A"
 complement(20119..20207)
 /rpt_family="MER103"
 complement(20561..20640)
 /rpt_family="MIR"
 21268..21544
 /rpt_family="AluSg"
 21545..21591
 /rpt_family="(CA)n"
 21610..21629
 /rpt_family="(TC)n"
 complement(21630..21898)
 /rpt_family="L1MB3"
 complement(21935..22327)
 /rpt_family="MLT11"
 complement(22354..22653)
 /rpt_family="AluSx"
 complement(22777..23084)
 /rpt_family="AluSg"
 complement(24611..24909)
 /rpt_family="AluJb"
 complement(24939..25247)
 /rpt_family="AluY"
 26278..26871
 /rpt_family="MER50"
 26872..27174
 /rpt_family="AluSg1"
 27175..27285
 /rpt_family="MER50"
 27287..27738
 /rpt_family="L2"
 27952..28149
 /rpt_family="MLT1J2"
 complement(28311..28462)
 /rpt_family="MLT1J2"
 complement(28847..29140)
 /rpt_family="AluSx"
 Query Match 7.1%; Score 140.6; DB 9; Length 87076;
 Best Local Similarity 45.4%; Pred. No. 2.9e-10;
 Matches 503; Conservative 0; Mismatches 604; Indels 0; Gaps 0;

OY 825 ACCGACTCCCTCAGTCAGCGCCGCTCAACACGCCCTGGGACGCCGGTATAGTAGTCTGGTC 885
DB 3919 ACCACACACACACACCTGCTACTTACACACACACACACACACACACACACACACACAC 3978
OY 886 GCGCGCGGACACAGCGCGGACACCTTACACCGTGGGTGACCGCGCGCGCGCGGACGAG 945
DB 3979 ACCAC 4038
OY 946 GTCATAACCTCGGTGCGTACGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1005
DB 4039 ATCACCAC 4098
OY 1006 CCGACCGCGGACGAG 1065
DB 4099 ACCAC 1125
OY 1066 CCGCGCGGAGCGGAG 4158
DB 4159 ATCACCAC 4218
OY 1126 GGAACAGCATGGCCACCGCGGCTTTCGGGGTGGGGGCTGCTCTCCAGGCGCCAC 1185
DB 4219 AC 4278
OY 1186 CCGAGCTGACCGCGGACAGAGTGAAGACCGCCCTCATCGAGAGCGCGGACATAGTCGCC 1245
DB 4279 ATCACCAC 4338
OY 1246 CCAAGGAGATAGCGGACATCGCTACGGTGGGTAGGGTGAAGTCTACAGAGCCATC 1305
DB 4339 ACCAC 4398
OY 1306 AAGTACGACGACTAGCGGACGCTACGTTGACGGGTGCGTGGCGGACAGGAGAGCGCC 1365
DB 4399 ACCAC 4458
OY 1366 ACCAC 1425
DB 4459 ACCAC 4518
OY 1426 GGTGAGGAGCATCGACCTTCTACTCTACGACCGGACGAGGAGGAGGAGGAGGAGGAG 1485
DB 4519 ACCAC 4578
OY 1486 TACACGCGCTACTAGGCTTGGAGAGGTCGGCTACTACACCGGACCGCGGAGACCTGG 1545
DB 4579 ACCAC 4638
OY 1546 AGGTGACAGGTCGTCAGCTACAGGGCGGGGCGGAGTACAGGTCGAGCTGTCAGGAC 1605
DB 4639 ACCACATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4698
OY 1606 GGGAGCCTCAGCAGTCCGCGGGGACCGGATCCAGACCGGACCGGACCGGACCGGAC 1665
DB 4699 ACCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 4758
OY 1666 CCGACACCGGACGAGAGCTTCAGCGGTTCCGTTAACGACTACTTGGGACACAGGCGAC 1725
DB 4759 ACCACCCCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4818
OY 1726 ACCTTCACCATGAGCTGACAGCGGTGCGACCAAGATACCGGTGACCTGACCTGACCT 1785
DB 4819 ACCAC 4878
OY 1786 ACTTCCTACAGCAGCTTCACTCTACTCTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1845
DB 4879 ACCAC 4938
OY 1846 TCGAGTGCAGCAGCAGCTACGAGCAGCGTCGAGTACCGGAGTACCGGAGTACCGGAGTAC 1905
DB 4939 GTCACACACATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4998
OY 1906 AGTTCTCCTCTACTACGCTTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1932

DB 4999 ACCAC 5025
RESULT 13
SCC24/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SCC24
Streptomyces coelicolor cosmid C24. 40899 bp DNA linear BCT 14-MAY-2002
AL163003 AL645882
AL163003.3 GI:20520830
acyl CoA carboxylase; alpha-L-arabinofuranosidase;
calcium-binding protein; carrier protein; fatty acid synthase;
gnt- family transcriptional regulator; integral membrane protein;
lipoprotein; protein kinase (regulator); rod shape-determining
protein; secreted peptidase; secreted protein; spermidine synthase;
sugar hydrolase; sugar kinase; secreted protein; spermidine synthase;
protein; sugar transporter; sugar binding protein; inner membrane
regulator; two-component sensor histidine kinase.
Streptomyces coelicolor A3(2).
Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 40899)
Redenbach, M., Kleser, H.M., Denapalte, D., Eichner, A., Cullum, J.,
Kinashi, H., and Hopwood, D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351
8843436
REFERENCE 2 (bases 1 to 40899)
Brown, S.P. and Harris, D.
Unpublished
3 (bases 1 to 40899)
Corden, A.M., Parzhill, J., Barrell, B.G. and Rajandream, N.A.
Submitted (03-APR-2000) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
On May 9, 2002 this sequence version replaced gl:7940255.
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Boehringer Ingelheim
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>) CDS are
numbered using the following system eg SC787.01c, SC (S.
coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the Frameplot program
of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nh.ox.jp/>
Jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, ggg, tgg or att) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
cloned clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid C24 lies

between and overlaps with cosmids C42 and 7A8 on the Asel-C genomic restriction fragment.

FEATURES
source
1. 40899
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid C24"
complement(1237..2742)
/gene="SCC24.02c", abfA,
/note="SCC24.02c" (2742)
complement(1237..2742)
/gene="SCC24.02c"
/note="SCC24.02c"
/note="SCC24.02c, abfA, alpha-L-arabinofuranosidase, len: 501 aa; almost identical to SW:ABFA_SRLI (EMBL:U04630) Streptomyces lividans alpha-L-arabinofuranosidase (EC 3.2.1.55) AbfA, 662 aa"
/codon_start=1
/transl_table=1
/product="alpha-L-arabinofuranosidase"
/protein_id="CAB86098.1"
/db_xref="GI:7414545"
/db_xref="SPTREMBL:Q9L0B5"
/translation="MRTAFTLDPAFTVGAVNPRFGSEVHLGRVYTGVPFGHPT ADAGLQDVLVRELGVAVRPGNFGVSGYKEDSVGPVEDPRRLDLARSTET NRGLSYIAFLKIGQPAEPMVNLTRGVAEALQELQYANHPSTGLDRSHG KDQFGLRLACLNEMGDMGTGHTAEYGRVAETARARQIDPDVEIVAGSSGQ SNETFAKSKKINLSIDENWYTKQAEVSALDNPAPRLLEDNYVMYAVVFGSL HYGARLRDRTVACLAOLYNIAPLMEFGSGPAMRQTTFEPESQSKYGRGVLQVR VDSPTIDAKYGRADLLHATVAREKSVTVFAVNSRTGALPELVALSGLELTVVE HSAIADAPQARMTLAEPRVVPVPHVDTSLRDLTALPELWSNLSRLR"
/complement(2770..3633)
/gene="SCC24.03c"
/note="SCC24.03c"
/complement(2770..3633)
/gene="SCC24.03c"
/note="SCC24.03c, possible sugar transporter membrane protein, len: 287 aa; similar to TR:CAB59597 (EMBL:AL132662) Streptomyces coelicolor probable sugar transport inner membrane protein SCF11.20, 299 aa; fasta scores: Opt: 673 z-score: 788.9 E(): 0; 35.3% identity in 278 aa overlap. Contains pfam match to entry PF00528 BPD_transp. Binding-protein-dependent transport systems inner membrane component and match to prosite entry PS00402 Binding-protein-dependent transport systems inner membrane comp sign. Contains also possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=1
/product="putative sugar transporter membrane protein"
/protein_id="CAB86097.1"
/db_xref="GI:7414546"
/db_xref="SPTREMBL:Q9L0B4"
/translation="MTTAAFPVTRPRNSVTPGQIVITLLGTAVSAVLAFLALVLT SKSSTEAVEPTHEKMDVDMGWAKALFEIGNIDMFVNSLYSVCTAYLVLSAL AGYGFARTEFGKGLMGVYAGLVSPAVIGVPLFTVQMGMDVTTVMFLPCAP AAWVILYKFPQGI:PRELEAPFIDAGRWGVFTIIVPLSRPSLAAGVITFEISSWN NEMLYMTVNNPDLTMPNGIATVNSYGIQWAGGLMAGGLIPLVYVFFVFORQIV AGVARTGLAGO"
complement(2902..3126)
/gene="SCC24.03c"
/note="PFam match to entry PF00528 BPD_transp. Binding-protein-dependent transport systems inner membrane component, score 81.20, E-value 2.2e-20"
complement(3037..3123)
/gene="SCC24.03c"
/note="PS00402 Binding-protein-dependent transport systems inner membrane comp sign."
complement(3630..4571)
/gene="SCC24.04c"
/note="SCC24.04c"
complement(3630..4571)

misc_feature
Query Match 7.1k; Score 140; DB 1; Length 40899;
Best Local Similarity 48.4k; Pred. No. 3.8e-10;
Matches 571; Conservative 0; Mismatches 570; Indels 39; Gaps 5;
QY 582 CGATGACCATGGACATGGAACATGCTTAGCTTCAATAGCAGCTGCTACTGGAGCAGCAG 641
DB 21445 CGACCGGACGCGCCAGCGACGACGCTCACCTCCACCGTGGCGGCGCGCGCCCTC 21386
QY 642 TAATGGCAAGTACAGAGGATGGCTCCAGGACGTAGCTGCGCGGAATTAGGCTTAGG 701
DB 21385 CGACGCAAGAGAGGGGTGTCCACACCCCGCGCGCGCTCGCGGAGGCTCTCGA 21326

/gene="SCC24.04c"
/note="SCC24.04c, possible sugar transporter membrane protein, len: 313 aa; similar to TR: CAB59596(EMBL:AL132662) Streptomyces coelicolor probable sugar transporter inner membrane protein SCF11.19, 316 aa; fasta scores: Opt: 533 z-score: 637.4 E(): 4.5e-28; 36.2% identity in 304 aa overlap. Contains pfam match to entry PF00528 BPD_transp. Binding-protein-dependent transport systems inner membrane component and match to prosite entry PS00402 Binding-protein-dependent transport systems inner membrane comp sign. Contains also possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=1
/product="putative sugar transporter membrane protein"
/protein_id="CAB86098.1"
/db_xref="GI:7414547"
/db_xref="SPTREMBL:Q9L0B3"
/translation="MTATSAETIVPARARTAAASGVRROFGHGHVAFPLALF ALFVYVPLRLGLYLSFTDANISGDSANFTIGDNTREALKDPLMDALGHSATFTILVY PCITVALLALLAHLEKRWLCFLPFLPSTVAANDNOMLFNPGTGNVHYV GLIEPLTKSYAMLAIVITLMTVGFSEFLYLAALOGIPAHLYEAALKDGANVHR ILHITLPMRLNITGLVVALQILASQVFDQAVVIOFGPGEDSTRTFVQYTLQGET SYRVGSAISIIIFVLIAAVALARMHLLRNREBGR"
complement(3641..3647)
/gene="SCC24.04c"
complement(3759..3986)
/note="PFam match to entry PF00528 BPD_transp. Binding-protein-dependent transport systems inner membrane component, score 39.00, E-value 1.1e-07"
complement(3837..3983)
/gene="SCC24.04c"
/note="PS00402 Binding-protein-dependent transport systems inner membrane comp sign."
complement(4568..5926)
/gene="SCC24.05c"
complement(4568..5926)
/note="SCC24.05c, possible sugar transporter sugar binding protein, len: 452 aa; similar to TR:CAB59595 (EMBL:AL132662) Streptomyces coelicolor, probable sugar transporter sugar binding lipoprotein SCF11.18, 442 aa; fasta scores: Opt: 274 z-score: 320.2 E(): 2.1e-10; 23.3% identity in 443 aa overlap. Contains pfam match to entry PF01547 SRP_bacterial.1. Bacterial extracellular solute-binding protein. Contains possible N-terminal signal peptide sequence"
/codon_start=1
/transl_table=1
/product="putative sugar transporter sugar binding protein"
/protein_id="CAB86099.1"
/db_xref="GI:7414548"
/db_xref="SPTREMBL:Q9L0B2"
/translation="MGRPLNRRHFLAAAGLTVAGSFCAALGTADALASGADTRY RYNLPSGSDGNMIALMDSPKDNPCIDVKSTLQNSPYTKLAAGAGNRPDLG VMHLGTFSPGRLLDPMOALLAKYGVREDDFNPDLKRAVLDGLKYLALPDIHQV LCFRVDVLLKAGLIGEDGRIVPVSETEENFDVLEARKATKAGLQITGLWHNDQNFQ

Query Match	6.8%	Score 134.8	DB 2	Length 169595
Best Local Similarity	46.5%	Pred. No. 1.6e-09		
Matches 470	Conservative 0	Mismatches 537	Indels 3	Gaps 1
814	AGCTCGGACGGACCGACTCCCTCAGTCAGCCGGCTCAACAACGGCTCGACCGCGGTATA	873		
DB	91287	ACCACCATCATCACCACCACTCATCACCATCACCACCAACCAACCAACCAACCAACCA	91228	
QY	874	GTAGTCTGGTGGCGCGGACAGCGGGCGCAACACACTACAGCTGGCTACCGCGCC	933	
DB	91227	ATCATCCTCCAGCATCATCATCAGCACCACTCACCACCAACCAACCAACCAACCAACCA	91168	
QY	934	GGCGCGAGGAGGTCATACCGTTCGGTGTGAGTTGACGACAGACACATCCGCACTTC	993	
DB	91167	ACCACCAACCACTCATCTACCATCACTACCACTGTCAACCAACCACTCACCAACCACT	91109	
QY	994	TCCAGCAGGGGACGCGCGGACGCGAGAGGCTCAAGCCGGAAGTCTGTCGCCCGCCGGGGT	1053	
DB	91107	ACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	91049	
QY	1054	GACATCATAGCCGCGCGGACCGGAAACAGCATGGCACCGCGATAAACAGCACTAC	1113	
DB	91047	ACGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	90988	
QY	1114	ACCAAGGCTCTGAGACGATGGGACCCCGCAGGTTGGGGGTTGGCGGCTCATC	1173	
DB	90987	ACCATCATCACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	90928	
QY	1174	CTCCAGGCCCAACCGAGTGGACCCCGGACAAAGGTGAAGACGGCCCTCATCGAGACCGCC	1233	
DB	90927	TCCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	90868	
QY	1234	GACATATGCCGCCCAAGAGATAGCGGACATCGCTACG---GTGCGGGTAGGGTGAAC	1290	
DB	90867	ACGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	90808	

```

RESULT 15
LOCUS      D83672
DEFINITION Streptomyces albobriscus DNA for subtilisin-like protease, complete cds.
ACCESSION D83672
VERSION   GI:1694626
KEYWORDS  subtilisin-like protease.
SOURCE    Streptomyces albobriscus (strain:S-3253) DNA, clone:SNM-P45.
ORGANISM  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (sites)
AUTHORS   Suzuki,M., Taguchi,S., Yamada,S., Kojima,S., Miura,K.I. and Momose,H.
TITLE     A novel member of the subtilisin-like protease family from Streptomyces albobriscus
JOURNAL   J. Bacteriol. 179 (2), 430-438 (1997)
MEDLINE   97144528
REFERENCE 2 (bases 1 to 3348)
AUTHORS   Taguchi,S.
JOURNAL   Unpublished
REFERENCE 3 (bases 1 to 3348)
AUTHORS   Taguchi,S.
TITLE     Direct Submission
JOURNAL   Submitted (22-FEB-1996) Seichi Taguchi, Science University of Tokyo, Dept. of Biological Science and Technology, 2641 Yamazaki, Noda, Chiba 278-8510, Japan (E-mail:staguchi@s.noda.sut.ac.jp, Tel:0471-24-1501(ex.4428), Fax:0471-25-1841)
COMMENT   Sequence updated (07-Nov-1996) by : Seichi Taguchi.
FEATURES  Location/Qualifiers
           1..3348
           /organism="Streptomyces albobriscus"
           /strain="S-3253"
           /db_xref="taxon:1897"

```

```
RBS
CDS
/clone="SAM-045"
2..7
14..3322
/Note="molecular masses observed on SDS/PAGE: 45kDa and 40kDa"
/codon_start=1
/transl_table=11
/product="subtilisin-like protease"
/protein_id="BAAL2040.1"
/db_xref="GI:1694627"
/translation="MRGDRHGRLVALAAGHAGPASAAGENSTAGSSAATALKGRH
RVLTLIGDRVALDAKGRVYGLPEAGREHPVQIRSDGHTLVVADARLAVASGLD
ORLPQVTELNKATRTAHRGKLVIVYRGAAAKADRVADGTVRTLTSLNADAVO
1PQAGAELEGEVTDGRTASGARVWLDQVRKASLDTSVQIGTPKAEAGYDGKGY
KIATVLDGVDATYHDLKQVYTSKNFTSAPTTGDVVGHGTHVASIAAGTGAOSKGYK
GVAPKGLINGVLDDAGDGDSDISLAGMWAAGADIVNMSLGGMDPTDPLEAA
VDLSAEKILFPAIAGNEGQSISPGSADSLTVGAVDDKDLADFSSTGPRLDG
AVKPDITAPGVDTAASAKNDIAKEVGEKPKAGYWTISGTSKATPHVAGAAALQQH
PEWYAEKGLUTASTKDGKYPTEQSGRVOVDKATQVIAEPVSLFGVQQHPHA
DDKPVTKLITYRLNLTEDVTLLKLTSTATGPKAPAGFTLGLASTLTPANGTASVD
VTADTRLGAVDGTYSAYVWATAGSVRTAAVEREVSINVTLLKLDPSKATANY
MAYLSGLTGLGDRSYPYEDGANVSRPKGYVLDASVLVGDPTWRGADHLAQP
KLDVTRNTTVVDARKAPKATVPKAAKQAFASADYTIETNDSAVSYGMWLENS
FRSAHLCPQITNGLSQMNTHFSNGAKAQYTAISGKVKKLATGYTRAFKAEFATV
OVGHAASAKKGAATVAPGHLPGSSGASGFSQEKLPSTRTLVLTNCGVTHDLDFEQ
LQVDNEGHPITYDAVYIGVYKTYKAGTYKETYNTAVFGPRITSSYGVFRDGNISIVG
V1PLFADGKHAGSSEFSSTVTLNKKKYSNNDPLFCGEGFTVPFGDAAYRLTTS
VRSKAAVAASTRIDASWTFRSKTSKTEKOLPVSSAREFAAVTGLDSKVAKKATFPV
VSCAAAGGKLNKSLAVTVSYNGKTKWTKTIVTKGLIIVKNPKAKAISFRKTIKDKG
NASLITIHNAVYK"
```

BASE COUNT 594 a 1234 c 1133 g 387 t

Query Match 6.5% Score 131.2; DB 1; Length 3348;
Best Local Similarity 52.0%; Pred. No. 8.4e-09;
Matches 435; Conservative 0; Mismatches 368; Indels 33; Gaps 5;

QY	412	CTGGATAGTCTGACGCTCAAGTTATGCGAACCTTAGCTTGGAACTTGGGATATGATGGT	471
DB	599	CTGGACACCTCGGTGCGGCAGATCGGCACCCGGAAGGCATCGGAGCGCGCTACGACGGC	658
QY	472	TCTGCAATCAATAGGAATATTCACACTGGAATTCAGCCTTCTCATCCAGATCTCCAA	531
DB	659	AAGGCGCTCAAGATCGCGCTCTGACACCGGTGTGAGACCGCCACCCCGGACCTCAAG	718
QY	532	CGAAAGTAATGGTGGGTAGATTGTTCATTCGTAGGAGTTATCCATACGATGACCAT	591
DB	719	GGCCAGGTGACCGGTCCAAAGAACTTCACCTCCGCGCCACCCACCGGTGACGTGGTC	775
QY	592	GGACATGGAATCATGTAGCTTCAATAGCAGCTGCTACTGGAGCAGCAAGTATGCAAG	651
DB	776	GGCCACGGCACCCACGCTGCGCTCCATCGCGCGGCGACGCGCCAGTCGAAAGGCGACG	835
QY	652	TACAAGGGAATGGCTCCAGGAGCTAAGCTGGCGGGAATTAAGTTCYTAGTCCGATGGT	711
DB	836	TACAAGGCGGTGCGCGCGCGGCGAAGATCCTCAGCGCAAGGTCCTCGACGACGCGCGG	895
QY	712	TCTGGAAGCATATCTACTATAATTAGGGAGTTGAGTGGCGCGTTGATACAAAGATAAG	771
DB	896	TTCCGCGAGCACTCGGGATCTCCGCGGGATGAGTGGCGGCGG-----CC	943
QY	772	TACGGAATTAAGGTCAATTAATCTTTCTTGTTCAGCCAGAGCTCCGACGGAACCCAC	831
DB	944	CAGGGCGCGGACATCGTCAACATGAGCCTCGCGGCGATGGACACCCCGGAGACCGCCG	1003
QY	832	TCCCTCAGTCAGCGCGCTCAACAGCGCTGGGAGCGCGGTATAGTAGTCTGCGTCCCGCC	891
DB	1004	CTGGAGCGCGGCTCGACAGCTCTCCGCGGAGAAGGGCATCTCTGTTCGCCATCGCGCG	1063
QY	892	GGCAACAGCGCGCGCAACACCTACACCGTGGCTCACCCCGCGCGCGGACGCAAGGTGATA	951
DB	1064	GGCAACAGCGCGCGCG-----CAGTCGATCGGTTCGCGCGGCGACGCGCCACTCCGCCCTC	1117

Search completed: November 2, 2002, 10:33:28
Job time : 3991.5 secs

QY	952	ACCGTCGGTCCAGTTGACAGCAACGACACATCGCCAGCTTCTCAGCAGCGGACCGAC-	1010
DB	1218	ACCGTCGGCGCGCTCGAGCAAGCAAGCTCGCGACTTCTCCTCCACCGCGCGCGC	1177
QY	1011	--CGCGAGCGGAAGGCTCAAGCCGGAAGTGGTGGCGCGCGCGGCTTGACATCATAGCGCG	1068
DB	1178	CTCGCGAGCGCGCGCGCTCAAGCCGCGACCTGACCGCGCGCGCGGTGGACATCAGCGCGCG	1237
QY	1069	CGCGCCAGCGGAACCAAGCATGGCCACCCCGATTAACGA-----CTACTACACCAAG	1119
DB	1238	TCGCGCAAGCGGCAACGACRTCCCAAGAGGTGGCGGAGAACCGCGCGCTACATGACC	1297
QY	1120	GCCTCTGGCAACGACGATGGCCACCCCGCACGTTTGGCGGCTTGCGCGGCTCATCTCCAG	1179
DB	1298	ATCTCGCGGACCTCGATGGCGCACCCCGCACGTCGCGCGCGCGCGGTGCTCAAGCAG	1357
QY	1180	GCCACCCCGAGCTGGACCCCGCGGACCAAGGTCAAGACCGCCCTCATCGAGACCGCGCA	1235
DB	1358	CAGCACCCGGAGTGGAACTAGCGCGGAGCTGAAGGGCGGCTCACCGCTCCACCAA	1413

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic acid search, using sw model

Run on: November 2, 2002, 08:10:53 ; Search time 1873.5 Seconds
(without alignments)
17090.193 Million cell updates/sec

Title: US-09-841-553-6
Perfect score: 1977
Sequence: 1 ATGAGGCGTGAAGCTCT.....AGGCGTCTACTACGGG 1977

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estlin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_estc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_est3:*
- 12: gb_est4:*
- 13: gb_est5:*
- 14: gb_estfun:*
- 15: em_estom:*
- 16: gb_gss:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mem:*
- 23: em_gss_mus:*
- 24: em_gss_oth:*
- 25: em_gss_pro:*
- 26: em_gss_rod:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93.4	4.7	641	BM181884	BM181884 fv51b11.y
2	86.4	4.4	681	13	CNS02600
3	83.2	4.2	500	17	B67199
4	81.6	4.1	440	10	BE637787
5	78.8	4.0	1116	11	AY110435
6	78.4	4.0	512	10	AV938606

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	BM181884	641 bp	mRNA	linear	EST 11-DEC-2001
BM181884	fv51b11.y1	Sugano SJD adult male	similar to contains element TARI	repetitive element	5'		
ACCESSION	BM181884	GI:17512842					
VERSION	BM181884						
KEYWORDS	EST						
SOURCE	zebrafish						
ORGANISM	Danio rerio						
REFERENCE	ACTINOPTERYGII; NEOPTERYGII; TELEOSTEI; CYPRINIFORMES						
AUTHORS	Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Pearson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.						
TITLE	WASHU Zebrafish EST Project 1998						
COMMENT	Unpublished (1998)						
	Contact: Stephen L. Johnson						
	Washington University School of Medicine						
	444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA						
	Tel: 314 286 1800						
	Fax: 314 286 1810						
	Email: zbrafish@watson.wustl.edu						

7	77.2	3.9	482	14	BQ170719	BQ170719	WHE1798.B
8	76.4	3.9	750	17	A2046831	A2046831	rhdb00880
9	76.2	3.9	705	14	BQ753023	BQ753023	WHE4122.A
10	75.8	3.8	422	13	B1098380	B1098380	WHE4122.A
11	75.6	3.8	645	14	BQ753508	BQ753508	WHE4122.A
12	75.4	3.8	773	17	CNS01VNC	CNS01VNC	WHE4122.A
13	75	3.8	627	12	BG608096	BG608096	WHE4122.A
14	73.6	3.7	559	10	A453236	A453236	WHE4122.A
15	73.6	3.7	807	14	BQ660503	BQ660503	WHE4122.A
16	73.6	3.7	849	13	B1641907	B1641907	WHE4122.A
17	73.6	3.7	692	14	BQ766288	BQ766288	WHE4122.A
18	73.6	3.7	852	12	BG310300	BG310300	WHE4122.A
19	73	3.7	585	10	BE470936	BE470936	WHE4122.A
20	73	3.7	617	10	BE361069	BE361069	WHE4122.A
21	72.8	3.7	627	13	BQ258334	BQ258334	WHE4122.A
22	72.8	3.7	645	9	A1389106	A1389106	WHE4122.A
23	72	3.6	877	13	B1952167	B1952167	WHE4122.A
24	71.6	3.6	443	13	BQ039802	BQ039802	WHE4122.A
25	71.6	3.6	651	13	BQ244245	BQ244245	WHE4122.A
26	71.2	3.6	915	12	BG343471	BG343471	WHE4122.A
27	71	3.6	534	10	BE495200	BE495200	WHE4122.A
28	71	3.6	608	13	BQ330059	BQ330059	WHE4122.A
29	70.8	3.6	511	13	BM004465	BM004465	WHE4122.A
30	70.2	3.6	820	13	B1956721	B1956721	WHE4122.A
31	70	3.5	659	13	BM371945	BM371945	WHE4122.A
32	69.4	3.5	545	12	BF727798	BF727798	WHE4122.A
33	69.4	3.5	551	13	BQ366220	BQ366220	WHE4122.A
34	69.2	3.5	612	10	BE195101	BE195101	WHE4122.A
35	68.8	3.5	317	13	BQ366180	BQ366180	WHE4122.A
36	68.6	3.5	733	10	AM587459	AM587459	WHE4122.A
37	68.4	3.5	494	13	BQ360881	BQ360881	WHE4122.A
38	68.2	3.4	970	17	CNS03H6V	CNS03H6V	WHE4122.A
39	67.8	3.4	690	12	BG810224	BG810224	WHE4122.A
40	67.8	3.4	701	14	BQ838521	BQ838521	WHE4122.A
41	67.4	3.4	677	13	BM569289	BM569289	WHE4122.A
42	67.4	3.4	722	14	BQ766040	BQ766040	WHE4122.A
43	66.8	3.4	536	10	BE603252	BE603252	WHE4122.A
44	66.5	3.4	527	10	AV952678	AV952678	WHE4122.A
45	66.4	3.4	570	14	BQ635733	BQ635733	WHE4122.A

Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center Clone
distribution information can be found through the I.M.A.G.E.
Consortium/LLNL, send email to: infoimage.llnl.gov
Seq primer: T3 ET from Amersham
High quality sequence stop: 527.
Location/Qualifiers

FEATURES
source

1. 641
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Sugano SUD adult male"
/sex="male"
/tissue_type="whole body"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notes="vector: pML85-FL3; Site 1: DraIII (CACCATGTC);
Site 2: DraIII (CACTGTGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTCCTTTTITTTTTTTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TGTGGCTAGCG], digested and cloned into distinct DraIII
sites of the pML85-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTC). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed and donated by Dr. Sumio
Sugano (University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTCTGCTTAAAGTCGCG and 3' end primer
CGACCTGCGAGTCGAGCACA."

BASE COUNT 218 a 234 c 37 g 152 t
ORIGIN

Query Match 4.7%; Score 93.4; DB 13; Length 641;
Best Local Similarity 47.7%; Pred. No. 3e-11;
Matches 274; Conservative 0; Mismatches 301; Indels 0; Gaps 0;

OY 1289 ACCTGACAGGAGCCATCAAGTAGAGAGTACGCGCAAGCTCACCGGCTCCGTCG 1348
DB 65 ACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 124
OY 1349 CCAGCAGGAGAGCCGACACACACCTGCGAGCTCAGCGCGCCACCTCTGTGACGCCA 1408
DB 125 ACTGCTACTGCGAGCTACATCACTTCCTCCACACACACACACACACACACAC 184
OY 1409 CCCTTACTTGGGACACGGGCTGAGGACATGACCTTACCTCTAGGACCCCGACGGGA 1468
DB 185 TCTCCACACACACACACACCTTCTCCACACACACACACACACACACACACAC 244
OY 1469 ACAGGTTGACTACTCTTACACCGCTTACTAGCGCTTCGAGAGGTCGGCTACTACACC 1528
DB 245 ACACACACACCTTCCAGCTACATCACTTCAACACACACACACACACACACAC 304
OY 1529 CGACCGCGGAGCTGGAGGCTGAAGTGTGAGCTACAGGCGGCGGCGGAGCTACCCAGG 1588
DB 305 TCTCCAGACACACACACACCTTCCAGCTACATCACTTCTCCACACACACACAC 364
OY 1589 TCGACGTCGTGACGCGGAGGCTTACGACGTCGCGGCGGCGGCGGCGGCGGCGG 1648
DB 365 CCAGCTACATCACTTCTCCACACACACACACACACACACACACACACACAC 424
OY 1649 CCACCGGAGACCCGACCCGACCCGACCCGACCCGACCCGACCCGACCCGACCC 1708
DB 425 ACACACACCTTCTCCAGCTACATCACTTCAACACACACACACACACACACAC 484
OY 1709 ACTGGGACACGACGACCTTCCAGTACGCTACAGGCGGCTGCGGCGGCGGCGGCGG 1768
DB 485 ACTTCAACACACACACACCTTCTCCAGCTACACCTTCTCCAGCAGCAGCAGCAGC 544
OY 1769 GTGACCTGACCTTCTGATATCTTCTACACGACCTTCCAGCTTCTGAGACCTTCTG 1828
DB 545 GCTACATCACTTCAACACACACACACCTTCTCCAGCTACATCACTTCAACACAC 604
OY 1829 GCAACCTCGTTGACAGGTCCTCCAGTCGAGCAACAGC 1863

DB 605 ACAAATCTCCAGCTACAACTTCTCCAGCAACAC 639
CNS02EOD 681 bp DNA linear GSS 13-MAY-2000
Tetraodon nigroviridis genome survey sequence 17 end of clone
262H14 of library G from Tetraodon nigroviridis, genomic survey
sequence.
RESULT 2
CNS02EOD/c
LOCUS
DEFINITION
ACCESSION AL193990.1 GI:7832096
VERSION GSS: genome survey sequence.
KEYWORDS Tetraodon nigroviridis.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
REFERENCE 1 (bases 1 to 681)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bertot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 681)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bertot, A. and
Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 681)
AUTHORS Direct Submission
TITLE Submitted (12-APR-2000)
JOURNAL This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
Location/Qualifiers
1. 681
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="G"
/clone_lib="G"
/note="genoscope sequence ID : C0AG262DB07LP1-end : T7"
BASE COUNT 74 a 105 c 170 g 282 t 50 others
ORIGIN
Query Match 4.4%; Score 85.4; DB 17; Length 681;
Best Local Similarity 45.8%; Pred. No. 1.3e-09;
Matches 190; Conservative 20; Mismatches 196; Indels 0; Gaps 0;

OY 1285 GTGAAGCTCTACAGGCGCATCAAGTACGAGCTACGCGGCGGCGGCGGCGGCGG 1344
DB 408 GTCCAC 349
OY 1345 GTCCGCGACAGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1404
DB 348 ACAAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 289
OY 1405 GGCACCCCTTCTAGTGGGACGAGGCTCGAGCGATCGACCTTCTAGCAGCCCAAC 1464
DB 288 GGCAGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 229
OY 1465 GGGACGAGGTTGACTACTCTTACACCCGCTTCTACGCTTTCGAGAGGTCGGTACTAC 1524
DB 228 AAC 169
OY 1525 AACCCGACCCCGGACCTCGAGGCTCAAGGTCGCTAGGTCGCTAGGCGGCGGCGGCGG 1584

/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/note="Vector: lambda Uni-ZAP XR, excised phagemid;
Site1: EcoRI; Site2: XhoI; plants were grown in the
greenhouse. Whole spkic with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give pBluescript
phagemids in the TJ Close lab (Choi, Close, Fencson) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."
BASE COUNT 77 a 163 c 140 g 60 t
ORIGIN

Query Match 4.1%; Score 81.6; DB 10; Length 440;
Best Local Similarity 54.3%; Pred. No. 1.6e-08;
Matches 190; Conservative 0; Mismatches 154; Indels 6; Gaps 1;
QY 979 AACATCCGAGCTTCTCCAGGAGGGAGCCGCGG-----ACGGAAGGCTCAAGCG 1032
DB 91 ATCGTGGCGCTTCTCTCGCGGGCGGCGAGCGGGCGAGCGCGCGTGTCTCAAGGCC 150
QY 1033 GAAGTCGTGCGCGCGGGTGTGACATATAGCCCGCGCGGCGAGCGGAGCCAGCATGGCG 1092
DB 151 GACATATGCGCGGGGCTCAACATCTCTCGCGGGTGGCGCTCGATGGTGGCCATCGAT 210
QY 1093 ACCCGGATAAAGACTACTACACCAAGGCTCTGGACACAGCATGGCGGCGGCGGCGGCGG 1152
DB 211 GCGAGGAGTGGTACAGCTTACACAGCTTGGCTCGCGGCGGCGGCGGCGGCGGCGGCGG 270
QY 1153 TGGGCGCTTGGCGGCTCTCTCCAGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1212
DB 271 ACCGCGATCTGAGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 330
QY 1213 ACCGCGCTCTGAGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1272
DB 331 TGGGCGATCTGAGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 390
QY 1273 GTGCGGCTGAGGCTTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1322
DB 391 GAGCAGCGGAGGAGGCTTCTACTCGATCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 440

RESULT 5
AY110435/c
LOCUS AY110435 1116 bp mRNA linear HTC 25-MAY-2002
DEFINITION Zea mays CL57831.1 mRNA sequence.
ACCESSION AY110435
VERSION AY110435.1 GI:21214844
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays.
REFERENCE
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.Y.
TITLE Maize Mapping Project/Dupont Consensus Sequences for Design of
Overgo Probes
JOURNAL Unpublished (2002)
AUTHORS 2 (bases 1 to 1116)
TITLE Coe,B.C.
REFERENCE Direct Submission
TITLE Submitted (25-APR-2002) Maize Mapping Project, University of
JOURNAL Missouri, Columbia, MO 65211, USA
FEATURES
source
1. .1116
/organism="Zea mays"
/db_xref="MaizeDB:632501"
/db_xref="taxon:4577"
/clone="CL57831.1"

/clone_lib="Maize Mapping Project/Dupont Consensus
Library"
/notes="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed Dupont contigs; this resource was
assembled by Dupont as part of a collaboration for the
overgo addressing of BACS in conjunction with the Maize
Mapping Project"
BASE COUNT 178 a 332 c 343 g 220 t 43 others
ORIGIN

Query Match 4.0%; Score 78.8; DB 11; Length 1116;
Best Local Similarity 49.0%; Pred. No. 9.4e-08;
Matches 194; Conservative 0; Mismatches 202; Indels 0; Gaps 0;
QY 1017 CGGAAGGCTCAAGCGGGAAGTCTGCTCCCGCGGGGTTGACATCATAGCCCGCGCGGCGAG 1076
DB 916 CGGAGTGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 857
QY 1077 CGGAACGACATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1136
DB 856 GGAGGTGCCCGTCTNNNNNNNNNAGTCGAGCAGCTTCAACGGCGGCTCTCCGCGGCGGCG 797
QY 1137 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1196
DB 796 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 737
QY 1197 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1256
DB 736 CACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 677
QY 1257 AGUGGACATCGGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1316
DB 676 CCAGATCATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 617
QY 1317 CTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1376
DB 616 TGTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 557
QY 1377 GAGCGTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1412
DB 556 CGGTACATCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 521

RESULT 6
AV938606
LOCUS AV938606 512 bp mRNA linear EST 18-JAN-2002
DEFINITION AV938606 K. Sato unpublished cDNA library, strain H602 adult,
heading stage top three leaves Hordeum vulgare subsp. spontaneum
cDNA clone bahl8j06 5', mRNA sequence.
ACCESSION AV938606
VERSION AV938606.1 GI:18234403
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. spontaneum.
ORGANISM Hordeum vulgare subsp. spontaneum
REFERENCE
AUTHORS Sato,K., Saitoh,D. and Takeda,K.
TITLE Barley EST sequencing project in NIG and Okayama Univ
JOURNAL Unpublished (2002)
COMMENT Contact: Tadashi Shin-i
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1. .512
/organism="Hordeum vulgare subsp. spontaneum"
/strain="H602"

/db_xref="taxon:77009"

/clone="bahl8106"

/clone_lib="K. Sato unpublished cDNA library, strain H602 adult, heading stage top three leaves"

/tissue_type="top three leaves"

/dev_stage="adult, heading stage"

86 a 202 c 150 g 73 t 1 others

Query Match 4.0% Score 78.4; DB 10; Length 512;
Best Local Similarity 49.2%; Pred. No. 9.2e-08;
Matches 239; Conservative 0; Mismatches 241; Indels 6; Gaps 1;

QY 1095 CCCGATTAACGACTACTACACAGGCTCTGGAACACAGCTGCGCACCGCACGCTTC 1154
DB 33 CCCGAGCGGCTCCGCTTCAACATCATCTCGGAGCTCATGTGCTGCGCGACGCTGAG 92
QY 1155 GGGCTTGGCGGCTCATCTCCAGGCGCCACCGAGCTGGACCGGACAGGTGAAGAC 1214
DB 93 CGGCTCGCGGCTCTCCGCGAGCGCACCGGAGCTGGAGCGGCGGCTCAAGTC 152
QY 1215 GCGCTCATCGAGCGCGACATAGTCGCGCCAGGAGATAGCGGACATCCCTACGG 1274
DB 153 GCGCTCATGACCGCGGTACA-----ACCTGACACATCTCCGGGAGATCATCAAGA 206
QY 1275 TCGGGTAGGTGAAGCTTACAGGCGCATCAAGTACGACGACTACGCCAAGCTCACCTT 1334
DB 207 CTGCGGACCGGAGCCAGTCGACGCGTTCGTCGCGCGCGCGGCGCACGTGACCCAA 266
QY 1335 CACCGCTCTCTCGCGCAAGGAGCGCCACCGACACCTTCAGCTAGCGGCGCAC 1394
DB 267 CAGCGGCTCAACCGCGGCTGTGTACGACCGGCGCACCGCGGCTTCCCT 325
QY 1395 CTTCGTGACCGCACCTTACTTGGGACAGCGGCTCGAGGACATCGACCTTACTCTA 1454
DB 327 CTGCGGCTGCTACATCTCATCCAGATCGCGCTTCACTAGGAGCGGCTCGCGTGC 306
QY 1455 CGACCCCAAGGAGAGGTGTACTCTTACACCGCTTACTAGCGCTTACGCTAGGAGGT 1514
DB 387 CGACTGCTCGAAGAGCGCGCGCTTCCGCGGACCTCACTACCGCGCTTCCGCGAGT 446
QY 1515 CGGCTACTAACCGGACCGCGCGGACCTGAGCGTCAAGCTGCTCAGTACAGGCGCG 1574
DB 447 CTTCGCTGCTACAGGACTCTCTCACTTACCAGCGGCTGTGAGCAAGCTCGGACGCA 506
QY 1575 GCGGAA 1580
DB 507 CCGGAA 512

RESULT 7
LOCUS B0170719
DEFINITION whe1798_b12_D242T wheat pre-anthesis spike cDNA library
ACCESSION B0170719
VERSION B0170719.1
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum

REFERENCE B0170719
AUTHORS Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Heia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L., and Tong, J.C.

TITLE The structure and function of the expressed portion of the wheat genomes - Pre-anthesis spike cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: andersn@pw.usda.gov

This EST was generated by sequencing from the 3' end of the clone. Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20.

Seq primer: T7 primer.

Location/Qualifiers

1. 482

/organism="Triticum aestivum"

/cultivar="Chinese Spring"

/db_xref="taxon:4565"

/clone="WHE1798_B12_D24"

/clone_lib="Wheat pre-anthesis spike cDNA library"

/tissue_type="Spike before anthesis"

/dev_stage="Adult plant"

/lab_host="g. coli SOLR"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;

Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give phagemids in the T7 close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 90 a 172 c 148 g 72 t

ORIGIN

Query Match 3.9%; Score 77.2; DB 14; Length 482;

Best Local Similarity 55.4%; Pred. No. 1.7e-07;

Matches 173; Conservative 0; Mismatches 133; Indels 6; Gaps 1;

QY 979 AACATGCGCAGCTTCTCCAGGAGCGGCGCGG-----ACGGAAGGCTCAAGCGG 1032

DB 141 ATCTGTGCGGCGCTTCTGCTCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 200

QY 1033 GAAGTGTGCGGCGGCGGCTTGACATCATAGCGCGCGCGCGCGCGGCGGCGGCGGCGG 1092

DB 201 GACATCATGCGCGGCGGCTCAACATCTTCCGCGGCGGCGGCGGCGGCGGCGGCGG 260

QY 1093 ACCCGCATTAACGACTACTACACAGGCGCTCTGGAACGACATGCGCCACCGCAGCGT 1152

DB 261 GGCACGAGCTGTACGCTACAACGTTTGCCTCCGCGACGTCCATGCGCGCGCGCGCG 320

QY 1153 TCGGCGGTTGGCGGCTCATCTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1212

DB 321 ACCGCGATCGTGGCGCTCGTGAAGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380

QY 1213 ACCGCGCTCATGAGCG 1272

DB 381 TCGGCGCATGACG 440

QY 1273 GGTGCGGGTAGG 1284

DB 441 GAGCAGCGAAGG 452

RESULT 8

LOCUS AZ046831

DEFINITION nbe0088017r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic

ACCESSION AZ046831

VERSION AZ046831.1

KEYWORDS GSS.

SOURCE Oryza sativa.

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

750 bp DNA linear GSS 16-MAR-2000

clone OSNBB0088017r, DNA sequence.

GI:7250600

Db 514 CACGTGAACCCGCGAGGCGCGTCCGACCCGCGCTCTAGCATCAGCCGCGGAC 573
 Qy 1369 CACACCTTCGAGTC 1383
 Db 574 TAGTCACCCACCTC 588

RESULT 9
 BQ753023 705 bp mRNA linear EST 23-JUL-2002
 LOCUS WHE4122_A06_A12S Wheat salt-stressed root cDNA library Triticum
 DEFINITION aestivum cDNA clone WHE4122_A06_A12, mRNA sequence.
 VERSION BQ753023
 KEYWORDS EST
 SOURCE BQ753023.1 GI:21930805
 ORGANISM bread wheat.
 Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Triticum.
 REFERENCE 1 (bases 1 to 705)
 AUTHORS Anderson, O.D., Akhunov, E., Chao, S., Crossman, C., Deal, K., Dvorak, J.,
 Lazo, G.R., Pham, J., Rausch, C.J., Wilson, C. and Woo, J.
 TITLE The structure and function of the expressed portion of the wheat
 JOURNAL genomes - Salt-stressed root cDNA library
 COMMENT Unpublished (2002)
 CONTACT: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov
 Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: SK primer
 FEATURES
 Location/Qualifiers
 1..705
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE4122_A06_A12"
 /clone_lib="wheat salt-stressed root cDNA library"
 /tissue_type="Roots"
 /dev_stage="Full tillering"
 /lab_host="E. coli SOUR"
 /note="vector: Lambda Uni-ZAP XR, excised phagemid
 plasmid SK(-); Site 1: EcoRI; Site 2: XhoI; Hydroponic
 plants grown to full tillering stage were treated with 150
 mM NaCl for either 12 hours or 7 days. Root tissues of the
 plants subjected to both types of treatment were collected
 separately at University of California, Davis (E. Akhunov
 and K. Deal in J. Dvorak's Lab). Total RNA was prepared
 separately from the two samples (12h and 7day treatments),
 and equal amount of RNA was then pooled. polyA RNA was
 purified from the pooled RNA, a cDNA library was made, and
 the cDNA clones were in vivo excised to give pBluescript
 SK(-) phagemids in J. Dvorak's lab (E. Akhunov, J. Dvorak)
 at the University of California, Davis. Colony plating,
 plasmid DNA preparations and DNA sequencing were performed
 in the OD Anderson lab (all other authors)."
 BASE COUNT 157 a 280 c 201 g 67 t
 ORIGIN

Query Match 3.9%; Score 76.2; DB 14; Length 705;
 Best Local Similarity 50.7%; Pred. No. 3.3e-07;
 Matches 265; Conservative 0; Mismatches 246; Indels 10; Gaps 3;
 Qy 1367 CCACACCTTCGAGTCAGCGCCACCTTCGTCGACCCGACCCCTCTACTGCGGACCG 1426
 Db 157 CCACCTTCGCGGCTCTCCCTCGCGACTCCATCTGTCGCCGCGACGCCGCCGCC 216

1 (bases 1 to 750)
 Wing, R.A. and Dean, R.A.
 A BAC End Sequencing Framework to Sequence the Rice Genome
 Unpublished (1998)
 CONTACT: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: GCTATTAGTGACATATAG
 Class: BAC ends
 High quality sequence stop: 339.
 Location/Qualifiers
 1..750
 /organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="OSJNB008017r"
 /clone_lib="CUGI Rice BAC Library (EcoRI)"
 /tissue_type="Leaf"
 /lab_host="E. coli DH10B"
 /note="Vector: pBACindigo; Site 1: EcoRI; Site 2: EcoRI;
 Rice is the most important food crop in the world. Half of
 the world population, especially those inhabiting highly
 populated areas of the humid tropics and subtropics, rely
 on rice as their primary source of carbohydrate.
 Monocotyledonous rice is a diploid plant (2n=24) with a
 haploid genome equivalent of 431 Mbp (Arumuganathan and
 Earle, 1991). The relatively small genome of rice, three
 times larger than that of Arabidopsis, makes it suitable
 for genomic studies. In order to facilitate positional
 cloning, physical mapping and genome sequencing of rice,
 we have constructed a BAC library from Oryza sativa.
 Nipponbare variety using EcoRI as the cloning enzyme. The
 library contains 55,296 clones with an average insert size
 of 121 kb providing approximately 15 haploid genome
 equivalents. The deep coverage allows the isolation of a
 particular sequence with a probability of 99.9%. Three
 high density filters, each containing 18,432 clones
 (doubly spotted), represent the whole library for colony
 screening and can be requested from the Clemson University
 BAC/EST Resource Center (www.genome.clemson.edu)."
 BASE COUNT 128 a 254 c 256 g 110 t 2 others
 ORIGIN

Query Match 3.9%; Score 76.4; DB 17; Length 750;
 Best Local Similarity 54.4%; Pred. No. 3e-07;
 Matches 204; Conservative 0; Mismatches 157; Indels 14; Gaps 2;
 Qy 1023 GTCACGCGGAGTGTGCGCCCGCGTGTGACATCATAGCCCGCGCGCGGAC 1082
 Db 214 GCTCAGCGCGAGTGTGCGCCCGCGGTGAACATCATCGCGCGCGGCAACT 273
 Qy 1083 CAGCATGGGACCCCGGATACGCA-----CTACTACACCAAGCCCTCTGGAAC 1130
 Db 274 GGGACCGTGGGGCTGGAGGCGAGCCCGCGGTCCGACCTCAGGTGCTCTGGGAC 333
 Qy 1131 CAGCATGGCACCACCGCGGTTGCGGGTTCGCGGCTCATCTCAGCGCCACCGAG 1190
 Db 334 ATCAATGGCGTGGCCGCGATCGCGGCGGCTCATCTAGTTCGCGGCGCCGCTC 393
 Qy 1191 CTGACCCCGGACAGGAGCCGCTCATCGAGAGCGCGGACATAGTCCGCCCAA 1250
 Db 394 GTGGAGCCCGCGGTGTGCGGTTCGCGATCATGACGCGCGGACGCGGCA 453
 Qy 1251 GGAGATGCGGACATCGCTACGCGGTGGCGGTGA--AGCTTCAAGGCCATCAG 1308
 Db 454 GGGCAAGCGGATCATGACGCGGAGCGGCGGAGCGCGGCGGCGGTACGCGTGGCGCGG 513
 Qy 1309 TACGAGGACTACGCAAGCTCACCTTCACCGGCTCGCGGACAGGAGCGGCAAC 1368

QY 1427 GCTCGAGGACATCGACCTTACCTCTACGACGCCCAACGGGAAACGAGGTGTGACTACTCCT 1486
 Db 217 CGCGCCCAACGCGGACCGGCGCTCTCTCCCTCCAGGCGGCCACAAACGCGG 276
 QY 1487 ACACCGCTTACTACGGCTTCGAGAGGTGCGGCTACTACAAACCGACCGCGGAACTGGA 1546
 Db 277 GAACGGATTCGCGCGCCCAAGATGAACGCGAGCTCAACGGGCGCGGCTCATCGGCT 336
 QY 1547 CGGTCAAGTTCGTACAGTACAGGCGCGGCGGAGCTACAGGTGCGACCTGTCAGGAGG 1606
 Db 337 CGGCAAGCTCGGCTTCGGC---CGACCAACACAGAGGCGCGGCTCAACAACG 393
 QY 1607 GGAGCTCAGCAGTTCGCGCGCGCGCAACCGGAATCCAAACCCCAA---CCGGAACCCA 1662
 Db 394 TCAGCTTCGCGCGCGCGCGCGCAACCGGAGCTTACGCAAGCGGCGGCTCAACAACG 453
 QY 1663 ACCCGACGACGACGACGACGACCTTCACCGGTTCCGTTAAGGACTACTGGGACACGAC 1722
 Db 454 GCCAAGCCCAACCAACCAATAACAGGCTTCGGGCTCAACAGATGGCTGCGCGGCG 513
 QY 1723 GACACCTTCACCA---TGACGCTCACAGGCGTGGCCACCAAGATACCGGTGACCTGACC 1779
 Db 514 GGGGCTACGCAACTTCAACGGCGGCGCAACGAGGGGTGAAGAGCTACTTCAACAAGTCC 573
 QY 1780 TTGATATCTTCTACAGGCTCGACCTCTACTCTACGACGCCCAACGCGACCTCGTT 1839
 Db 574 ATCGGAGGCGCGCGCGCAACCAACAGCAGCAGCAACACTTCAGCGGCTACGGCGCG 633
 QY 1840 GACAGTTCACGTCGACGACACCTACGAGGAGCTCGAGTACG 1882
 Db 634 AAGAGGGCGCGCGGAGCGCAAGAGAGACGACGCCAAGAACG 676

RESULT 10

BIO98380
 LOCUS
 DEFINITION
 mRNA sequence.

ACCESSION
 BIO98380.1
 VERSION
 BIO98380.1
 KEYWORDS
 EST.

ORGANISM

sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Sorghum.

REFERENCE
 1 (bases 1 to 422)
 Klein,R.R., Cordonnier-Pratt,M.-W., Gingle,A., Sudman,M. and Pratt
 L.H.

TITLE
 An EST database from Sorghum: developing preanthesis pannicles

JOURNAL
 COMMENT
 Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for high quality sequence is
 20. Three-prime sequences, which are obtained with PolyTmix or T7
 sequencing primer, are presented as the reverse complement.
 Seq primer: JEN REV

High quality sequence stop: 419

POLYA-NO.

FEATURES

source

1. .422

Location/Qualifiers

/organism="Sorghum bicolor"

/cultivar="BPM623"

/db_xref="taxon:4558"

/clone_lib="Immature pannicle 1 (IP1)"

/note="Organ: Developing preanthesis pannicles; Vector:

plasmid pSK(-) from Lambda Zap II; Site_1: XhoI;

Site_2: EcoRI; The library was made from poly-A RNA in the
 cloning vector lambda ZAP II. Clones to be sequenced were
 prepared by mass excision.

BASE COUNT 75 a 155 c 135 g 57 t
 ORIGIN

Query Match 3.8%; Score 75.8; DB 13; Length 422;
 Best Local Similarity 54.5%; Pred. No. 3.5e-07;
 Matches 152; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 1018 GGAAGCTCAAGCGGAGTCGCGCGCGCGGCTTGACATCATAGCCCGCGGCCACG 1077
 Db 114 GCGTCTGAGCGCGGAGCTCATGGCGCGCGGCTCAACCTCTCGCGGCTGCGTGG 173

QY 1078 GGAACAGATGGGACCCCGGATAACGACTACTACACCAAGGCGCTCTGMAACGACATG 1137
 Db 174 GAGGTGCGCGCTCGGGGCGCGGAGTCGAACAGCTTCAACGCTCATCTCCGCGAGTCCNTG 233

QY 1138 GCCACCGCGGACGTTTCGGCGGCTTCGCGCTCATCTCCAGGCGCCACCGAGCTGGACC 1197
 Db 234 GCGAGCGCGACATCACCGGCTGCGGCTGCGGCTGCAAGAAGCGGACCGGACTGCTCG 293

QY 1198 CCGGACAAGGTGAAGACCGCCCTCATCGAGACCGCGGACATAGTCGCCCGCCCAAGGAGATA 1257
 Db 294 CCGCGCGGATCAAGTTCGCGCTCATGACCGCTCCAGCGCGCTCGACACACCGCGCAAC 353

QY 1258 GCGGACATCGGCTACGCTGCGGCTAGGCTGACGCTCTAC 1296
 Db 354 CAGATCATGGACGAGGAGCAGCGGAGCGGAGCTTCTAC 392

RESULT 11

BIO753506

LOCUS

DEFINITION

mRNA sequence.

ACCESSION

BIO753506.1

VERSION

BIO753506.1

KEYWORDS

EST.

SOURCE

ORGANISM

Sorghum vulgare.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 1 (bases 1 to 645)

REFERENCE

1 (bases 1 to 645)

AUTHORS

Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
 Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.

TITLE

Development of barley Transcriptome Resources

JOURNAL

Unpublished (2001)

COMMENT

Contact: Haugh R, Marshall DF
 Genome Dynamics/Computational Biology
 Scottish Crop Research Institute
 Invergowrie, Dundee, DD2 5DA, Scotland, UK
 Tel: 00 44 1382 562731
 Fax: 00 44 1382 562426
 Email: est@scri.sari.ac.uk.

FEATURES

source

1. 645

Location/Qualifiers

/organism="Sorghum vulgare"

/cultivar="Optic"

/db_xref="taxon:4513"

/clone_lib="EBan01_SQ004_H09"

/clone_lib="anther, yellow stage, no treatment, cv Optic,
 EBan01"

/tissue_type="anther"

/dev_stage="yellow stage"

/lab_host="DH108"

/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
 Non-normalised library, directionally cloned into pSPORT1.
 Derived from anthers dissected from developing flowers in
 glasshouse grown barley plants. Developed as part of the
 barley transcriptome resources of BBSRC/NERAC funded

```

cereal IGF (Investigating Gene Function) project.
BASE COUNT 124 a 202 c 233 g 86 t
ORIGIN
Query Match 3.8%; Score 75.6; DB 14; Length 645;
Best Local Similarity 55.2%; Pred. No. 4.5e-07;
Matches 169; Conservative 0; Mismatches 134; Indels 3; Gaps 1;

QY 1106 ACTACTACACCAAGGCTCTGGAAACAGACAGTGGCCAGCCGACGCTTTCGGCGGTGGCG 1165
DB 14 ACAGCTACACAGTGTCTCGGACAGCTGCTATGCGACCGCCGACGCTACCGGCGTGG 73
QY 1166 CGCTATCTCTAGGCGCCACCGAGCTGACCCCGGACAGGTGAGAGCCGCCCTATCG 1225
DB 74 CACTGTGAGAGGCGCCACCGGAGTGTGCGCTGCGGCGTCAAGTCCGCCATCATGA 133
QY 1226 AGACCCCGACATAGTGCCTCCCAAGGAGATAGCGGACATGCGCTACGGTGGGTAGGG 1285
DB 134 CCAGCTCCAGCACCCCGACAGCGGGGAACCCATCATGAGAGGAGGACCGGAGG 193
QY 1296 TGAAGCTCTAC---AAGGCTATCAAGTACGACGACTAGCCAGCTCACCTTCACCGGCT 1342
DB 194 CGAGCTTCTACTGATCGGCGCGGCGACGCTGATGCGCCAGGCGGTGGACCCGGTC 253
QY 1343 CCGTCCGCCACAGGAGGAGCCACCCACACCTTCGAGCTACGCGGCGGCGACCTTCGTGA 1402
DB 254 TGGTCTACGACCTCGGCGTGGCGGACTACAGCGGCTATCATCTGCGGCGCTGCTCGGCGAGG 313
QY 1403 CCGCCA 1408
DB 314 CGGCCA 319

RESULT 12
CNSOLVTC 773 bp DNA linear GSS 12-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence 77 and of clone 199C24 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL169549.1 GI:7807606
VERSION GSS; genome survey sequence.
KEYWORDS Tetraodon nigroviridis, Tetraodon nigroviridis, Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 773) Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F., Saurin, W. and Weissenbach, J. Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 773) Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
TITLE Unpublished
JOURNAL 3 (bases 1 to 773) Genome. Submitted (12-APR-2000)
AUTHORS This sequence is a single read and was generated as part of a large scale clone-end and sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1..773 /organism="Tetraodon nigroviridis"

```

```

/db_xref="taxon:99883"
/clone="199C24"
/clone_lib="c"
/note="Genoscope sequence ID : COAG199BB12LPI-end : T7"
BASE COUNT 210 a 265 c 69 g 184 t 45 others
ORIGIN
Query Match 3.8%; Score 75.4; DB 17; Length 773;
Best Local Similarity 43.1%; Pred. No. 5.3e-07;
Matches 259; Conservative 18; Mismatches 324; Indels 0; Gaps 0;

QY 1374 CTTCGAGCTCAGCGGGCCACCTTCGTGACCGCCACCTCTACTGGACACAGGGCGCTCGAG 1433
DB 166 CTGCTACTACTACWACTGCTACWACTAGTCWACWACWCTACTGCTACTACTACTACTA 225
QY 1434 CGACTCGACTCTACCTCTAGGACCCACGACGAGGAGGTGACTACTCTCTACTACTACCG 1493
DB 226 CTAGTGTACTACTGCTACTGCTACTGCTACTACTACTACTACTACTACTACTACTACTA 285
QY 1494 CTACTAGGCTTGGAGAGGTGGGCTACTACACCGACCGCCCGGAGACCTGGAGGTCAC 1553
DB 286 CTACWACTGCTACWACTGCTACWACTGCTACTACTACTACTACTACTACTACTACTACTA 345
QY 1554 GGTGCTCAGCTACAGCGGGCGGGGAACTACACGCTCGAGCTGTCAGCGACGGAGCCT 1613
DB 346 CTACTCTACTACTACTGCTACWACGAGCATACATACTACTGCTACTACTACTACTACTA 405
QY 1614 CAGCCAGTCCGGGGGGGACCCCGATCCAAACCCACCGACCCCAACCCCGACCCAC 1673
DB 406 CTGCAACTACWCACTGCTACWCTCTACGCACTACTACTACTACTACTACTACTACTACTA 465
QY 1674 CGACACCCAGACCTTCACCGGTTCCGTTAAGACTACTGGGACACCGACACCTTCAC 1733
DB 466 CWRCACTACTGCTACCACTACTCTCWACTGCTACTACTACTACTACTACTACTACTACTA 525
QY 1734 CATGAGCTCAGACGCTGGCGGACCAAGATAACGGTGCACCTGACCTTCGAACTCTCTA 1793
DB 526 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTG 585
QY 1794 CAACGACCTCGACTCTACTCTAGGACCCCGACCGCACTCTCTGTTGACAGGTCCAGTC 1853
DB 586 CTACWACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 645
QY 1854 GAGCAACAGCTAGGACAGCTGAGTACGCCAACCCCGCCCGGGAACCTGGAGCTTCCT 1913
DB 646 AACATACTACTGCTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTA 705
QY 1914 CGTCTACGCTCAGACACCTACGGCTGGGGGAGTACCGAGCTCAAGCGGCTCTACTA 1973
DB 706 CTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTG 765
QY 1974 C 1974
DB 766 C 766

RESULT 13
EG608096 627 bp mRNA linear EST 17-APR-2001
LOCUS WHE2496_H02_004Z5 Triticum monococcum early reproductive apex cDNA library Triticum monococcum cDNA clone WHE2496_H02_004, mRNA sequence.
ACCESSION EG608096
VERSION EG608096.1 GI:13658079
KEYWORDS EST.
SOURCE Triticum monococcum.
ORGANISM Triticum monococcum. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 627) Anderson, O.D., Chao, S., Dubcovsky, J., Echenique, V., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L.,

```

TITLE
Stamova, B. and Tong, J.C.
The structure and function of the expressed portion of the wheat
genomes - Early reproductive apex cDNA library from Triticum
monococcum
JOURNAL
Unpublished (2001)
COMMENT
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@wusda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Strategene SK primer.
Location/Qualifiers
1. 627

FEATURES
source

/organism="Triticum monococcum"
/cultivar="DV92"
/db_xref="taxon:4568"
/clone="WHE2496.H02.004"
/clone.lib="Triticum monococcum early reproductive apex
cDNA library"
/tissue_type="Early reproductive apex"
/dev_stage="Seven week-old plants"
/lab_host="E. coli XL0LR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; The tissue, total RNA, and
poly(A) RNA were prepared from apex at double-ridge stage
to terminal-spikelet stage during transition from
vegetative state to flower state, a cDNA library was made,
and the cDNA clones were in vivo excised at the
University of California, Davis (V. Echenique, B. Stamova
, J. Dubcovsky). Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."
BASE COUNT 142 a 247 c 173 g 55 t
ORIGIN

Query Match 3.8%; Score 75; DB 12; Length 627;
Best Local Similarity 49.3%; Pred. No. 6.1e-07;
Matches 258; Conservative 0; Mismatches 255; Indels 10; Gaps 2;
QY 1367 CCCACACCTTCGAGCTCAGCGGCGGACCTTCGTGACCGCCACCCCTCTACTGGGACAGG 1426
DB 76 CCAGCTTCCTCAGCCTCTCCTCGCGACTCATCTGTCGCCGCGGACGCGCGCCGCC 135
QY 1427 GCTCAGCGAGATCGACCTCTACTCTAGACCCCAACGGGACGAGGTGTGACTACTCT 1486
DB 136 GGGCCGCAAGCGGACCGGCGCTCTCTCCCTCCAGCGCGGCCCAACACGCGG 195
QY 1487 ACACCGCCTACTAGCGGTCGAGAGGTGGCTACTACACCGACCGCGGACCTTGA 1546
DB 196 GAACCGGATTCGGCGGCGGCAAGATGAACGCTAGCCTCAACGCGCGCGGCTCACTCGG 255
QY 1547 CGGTCAAGGTCTCAGCTACAAGGGGCGGGAACCTACAGGTCTGCTGTCAGCGAGC 1606
DB 256 CGCGCAAGCTCGGCTTCGGC---GCCACCAACACGAGCGGCGGCGGCGGCGGCG 312
QY 1607 GGAGCGCTCAGCGAGTCCGCGGCGGCAACCGGATTCACACCCCAACCGCAACCC 1666
DB 313 TCAGCTCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 372
QY 1667 CGACCGCGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1726
DB 373 CCAAGCCCAACACCAATACACGCGTTCGGGCTCAACAGATGACTCGCGCGCGG 432
QY 1727 CCTTCAACATGACGTCAACAGCG-----GTCCCAACCAAGATAACCGGTGACTCAAC 1779
DB 433 GGGGCTACACGAACCTTCAACCGCGGCGGCAACGAGGGGGGTCAAGAGTACTTCAACAGTCC 492
QY 1780 TTCGATACTTCCCTACACGACCTCGACCTCTACCTCTACGCGCGGCGGCGGCGGCGT 1839

DB 493 ATCGGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 552
QY 1840 CACAGGTTCACGTTCGAGGACGACGTCAGGACGACGTCAGTACG 1882
DB 553 AAGAAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 595

RESULT 14
AW453236
LOCUS

DEFINITION
AW453236
66003F06.y1 660 - Mixed stages of anther and pollen Zea mays cDNA,
mrna sequence.

ACCESSION
AW453236
VERSION
AW453236.1 GI:6994022

KEYWORDS
EST
SOURCE
Zea mays.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE
AUTHORS
TITLE

Maize ESTs from various cDNA libraries sequenced at Stanford
University

JOURNAL
Unpublished (1999)

COMMENT
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave. Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660033 row: F column: 06.
Location/Qualifiers
1. 559

FEATURES
source

/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone.lib="660 - Mixed stages of anther and pollen"
/tissue_type="whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="XL0LR"
/note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
Site_2: XhoI; Anther and pollen cDNA library.
Directionally sequenced with 5' end at the EcoRI site.
Created by Amie Franklin."
BASE COUNT 106 a 198 c 177 g 78 t
ORIGIN

Query Match 3.7%; Score 73.6; DB 10; Length 559;
Best Local Similarity 50.0%; Pred. No. 1.3e-06;
Matches 184; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 1017 CGGAGGCTCAGCGGCGGAGTCTCGCCCGCGGCGTTCACATCATAGCCCGCGGCGGCGG 1076

DB 192 CGAGTCTGAAGCGCGGACATCATCGCGCGGCTCAACATATCTCGCGCGCTGCGCGCTC 251

QY 1077 CGGACACGATGGGACCGCCCGGCGGCTTACACCAAGGCTCTGGAACACGAT 1136

DB 252 CGAGTGCGGCTCGGGGCGGCGGCGGCTTCAGCGGCTCTCGGCGGCGGCGGCGGCGG 311

QY 1137 GCGGACCGCGGCGGCTTCGCGGCGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1196

DB 312 GCGGCGGCGGATATACCGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 371

QY 1197 CCGGACAGGTGAAGACCGCGGCTTCATGACAGCGCGGCGGCGGCGGCGGCGGCGGCGG 1256

DB 372 CACCGCGGATCAAGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 431

QY 1257 ACCGACATCGGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1316

DB 432 CCAGATCATGGACGAGGAGGACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 491

QY 1317 CTAGCGCAAGCTTACCTTCACCGGCTCGGTGCGCCGACAGGGAAGCGCCACCCACACCTT 1376
 DB 492 TGTCCTCGGCAAGCGGTAGACCCCGGCTGTGTGACGACCTCGGCGTCCACGACTACGC 551
 QY 1377 CGACGTCA 1384
 DB 552 CGGCTACA 559

RESULT 15
 B0660503
 LOCUS H102008u HI Hordeum vulgare cDNA clone HI02005 3-PRIME, mRNA
 DEFINITION sequence.
 ACCSSION B0660503
 VERSION B0660503.1 GI:21801636
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.
 1 (bases 1 to 607)
 Michael, W., Wesche, W., Pleissner, K.-P. and Graner, A.
 EST sequencing and analysis in barley
 Unpublished (2000)
 CONTACT: Stein Nils
 Molecular Markers Group, Department Genbank
 Institute of Plant Genetics and Crop Plant Research (IPK)
 Corrensstr. 3, 06466, Gatersleben, Germany
 Tel: 039482-5522
 Fax: 039482-5595
 Email: stein@ipk-gatersleben.de
 Insert Length: 607 Std Error: 0.00
 Plate: 2 row: 0 column: 5
 Seq primer: M13uni.
 Location/Qualifiers
 1..607
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="GABI:136895"
 /db_xref="taxon:4513"
 /clone="HI02005"
 /clone_lib="Hi"
 /tissue_type="female inflorescences"
 /dev_stage="female inflorescences (approx. 3 mm in size)"
 /lab_host="XL10-Gold"
 /note="vector: pBluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning, to excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: ALSO due to the cloning system used Blue/white selection for recombinants is not 100% reliable."
 132 a 251 c 162 g 62 t

BASE COUNT 132 a 251 c 162 g 62 t
 ORIGIN

Query Match 3.7%; Score 73.6; DB 14; Length 607;
 Best Local Similarity 48.49; Pred No. 1.3e-06;
 Matches 236; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

QY 1397 TCGTGACCGCCACCTCTACTTGGGACACGGGCTCGAGGACGACATCGAACCTTACCTTACG 1456
 DB 120 TCCAGAGCGGCGAGCTTCTCGCGGCTCTCCCTCGGCGACTCCATCTGTGTCCTCCCGCCCG 179
 QY 1457 ACCCCACGGGACGAGGTGACTTACTCTACACGCGCTTACTACGCTTCGAGAGGTG 1516
 DB 180 ACAACACGCCCCCGCGCGGACGCGGCGCGGCTCTTCTCCCTCCCGCGCG 239
 QY 1517 GCTACTACACCGCGCGCGGACCTCGGACGCTCAGGTGCTGCTACGCTACACCGCGCG 1576
 DB 240 CCGCCCAACAC---AACGCGGCTTCCAAAGCCCAACACAGCCTCAACGCGCGCGCGCTCA 296

QY 1577 CGAACTACAGGTCTGACGCTCGTACGCGACGGGAGGCTTCAGCCAGTCCGGGCGGCAAC 1636
 DB 297 TCGGTCGCGCAAGCTCGCTTCGCGCCACACACCAAGGCGGCGCTTACAACA 356
 QY 1637 CGAATCCAAACCCCAACCCGGAACCCGACCCGACCGACCGACCGACCGACCTTACCGGTT 1696
 DB 357 ACOTCAGCCTCCCGACGCGCCGCTACGCCAACGCCAACGCCAACGCCGCGCTACGGCG 416
 QY 1697 CCGTTAAGCACTACTGGGACACGACGACCTTACCATGATGAACGCTAACAGCGCTGCCA 1756
 DB 417 CCAAGCCCAACAAACGCAACATCAACGCGCAACACCTTCGGGCTCAGCAAGATGCGCCCA 476
 QY 1757 CCAAGATAACCGGTGACCTGACCTTCGATATCTTCTTACAAACGACCTCGACCTCTACCTCT 1816
 DB 477 GCGCGGCGCGGCTTACAGCAACTTCAACGCGCGCAACGCGGCGTCAAGAGCTTACTTCA 536
 QY 1817 ACAGCCCAACCGCAACTCTGTTGACAGGTTCACGCTCGAGGCAACAGCTACGAGCACCTCG 1876
 DB 537 ACAAGTCCATCGGCGGCGCGGCGCAACAAACAGCAACAGCAACGCAACTTCAACTTCAACG 596
 QY 1877 AGTACGCC 1884
 DB 597 GCTACGGC 604

Search completed: November 2, 2002, 11:24:45
 Job time : 1887.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 2, 2002, 00:34:13 : Search time 281 Seconds
(without alignments)
15844.131 Million cell updates/sec

Title: US-09-841-553-6
Perfect score: 1977
Sequence: 1 ATGAAGGCGCTGAAGCTCT.....AGGCCGTCGTACTACGGG 1977

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185139 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq.101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1977	100.0	1977	18	Protease coding se
2	1366.4	69.1	1962	18	Pyrococcus furiosu
3	1366.4	69.1	1962	20	Hyperthermostable
4	1360.8	68.8	1977	18	Thermococcus prote
5	1360.8	68.8	1977	20	WO9856926 Seq ID 1
6	1165	58.9	1859	18	Thermococcus prote
7	969.4	49.0	1566	18	Pyrococcus furiosu
8	837	42.3	1235	20	Hyperthermostable
9	669.8	33.9	698	17	Hyperthermostable

10	507	25.6	507	18	AAV03761	Pyrococcus proteas
11	427	21.6	564	17	AAV08134	DNA sequence. Pyr
12	427	21.6	564	18	AAV85576	Thermococcus prote
13	341.6	17.3	419	18	AAV85691	Pyrococcus proteas
14	197	10.0	1461	18	AAV85679	Thermococcus prote
15	135	9.9	2121	23	ABL34900	T. yonsei subtile
16	132	7.7	546	23	ABL53453	T. yonsei subtile
17	131	6.6	2539	18	AAV61454	Streptomyces virid
18	131	6.6	2809	18	AAV61455	DhpA-mel chimeric
19	115.4	5.8	985	24	ABQ44816	Oligonucleotide fo
20	115.4	5.8	985	24	ABQ44817	Oligonucleotide fo
21	107.6	5.4	5163	19	AAV20700	Cryptosporidium pa
22	107.6	5.4	5163	21	AAV61849	ORF encoding a por
23	107.6	5.4	5163	24	AAV04778	C. parvum GP900 gen
24	107.6	5.4	5318	19	AAV20701	C. parvum GP900 gen
25	107.6	5.4	5318	21	AAV61848	DNA encoding a por
26	107.6	5.4	5318	24	AAV04777	HPV16-L2 encoding
27	106	5.4	4422	24	ABL59978	Human breast cell
28	104.6	5.3	1635	22	ABA49946	Human foetal liver
29	104.6	5.3	1635	22	ADA67865	Human foetal liver
30	104.6	5.3	1635	22	ABA34921	Probe #13387 for g
31	104.6	5.3	1635	22	AAK16270	Human brain expres
32	104.6	5.3	1635	22	AAK16270	Human bone marrow
33	104.6	5.3	1635	22	AAI22780	Probe #12713 for g
34	104.6	5.3	1635	22	AAI45082	Probe #16786 used t
35	104.6	5.3	1635	22	AAI08454	Human genome-deriv
36	104.6	5.3	1635	22	AB516047	Human breast cell
37	104.6	5.3	1635	22	ABA55261	Human foetal liver
38	104.6	5.3	1635	22	ABA25005	Probe #3471 for ge
39	104.6	5.3	1635	22	AAK03514	Human brain expres
40	104.6	5.3	1635	22	AAK28970	Human bone marrow
41	104.6	5.3	1635	22	AAI13556	Probe #3489 for ge
42	104.6	5.3	1635	22	AAI34918	Probe #1604 used t
43	104.6	5.3	1635	22	AAI03446	Probe #3437 used t
44	104.6	5.3	1635	22	AAI03446	Human genome-deriv
45	104.6	5.3	1635	24	AB503504	

ALIGNMENTS

RESULT 1
AAT85669
ID AAT85669 standard: DNA: 1977 BP.
XX
AC AAT85669;
XX
DT 20-APR-1998 (first entry)
XX
DE Protease coding sequence.
XX
KW Protease; research reagent; thermal stability; ss.
XX
OS Synthetic.
XX
PN WO9721823-A1.
XX
PD 19-JUN-1997.
XX
PF 07-NOV-1995; 96WO-JP03253.
XX
PR 12-DEC-1995; 95JP-0323285.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Asada K, Kato I, Mita M, Norishita M, Takakura H;
PI Tsunawasa S, Yamamoto K,
PI WPI: 1997-332794/30.
PI P-PSDB: AAW24123.
XX
PT Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are

PT useful industrially and as research reagents

PS Claim 11; Page 95-97; 159pp; Japanese.

XX This sequence represents the coding sequence for a protease of the
CC invention. The proteases of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries.

XX Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 other;

Query Match 100.0%; Score 1977; DB 18; Length 1977;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	841	CAGGCGCTCAACAAACGCTTGGAGCGCGGTATAGTAGTCTGCGTGGCGCGCAACAGC	900
Qy	901	GGGCGGAACACCTACACCGTGGGTACCGCGCGCGCGCGCGCGGTATACCGTGGT	950
Db	901	GGGCGGAACACCTACACCGTGGGTACCGCGCGCGCGCGCGGTATACCGTGGT	950
Qy	961	CGAGTTGACAGCAACGACCAACATCGCCAGCTTCTCCAGCAGGGGACCGCGCGGGA	1020
Db	961	CGAGTTGACAGCAACGACCAACATCGCCAGCTTCTCCAGCAGGGGACCGCGCGGGA	1020
Qy	1021	AGGCTCAAGCGGGAAGTGTGCGCGCGCGGTTCACATCAATAGCCCGCGCGCGGGA	1080
Db	1021	AGGCTCAAGCGGGAAGTGTGCGCGCGCGGTTCACATCAATAGCCCGCGCGCGGGA	1080
Qy	1081	ACCAGCTGGGCAACCGCGGTAAACGACTACTACACCAAGGCGCTCTGGAACAGCATGCC	1140
Db	1081	ACCAGCTGGGCAACCGCGGTAAACGACTACTACACCAAGGCGCTCTGGAACAGCATGCC	1140
Qy	1141	ACCGCGGACGTTTGGCGGCTTGGCGGCTCATCTCCAGCGCCACCGAGCTGGACCCCG	1200
Db	1141	ACCGCGGACGTTTGGCGGCTTGGCGGCTCATCTCCAGCGCCACCGAGCTGGACCCCG	1200
Qy	1201	GACAAAGTGAAGACCGCGCTCATTCGAGACCGCGGACATATCGCCCGCGAAGGATAGCG	1260
Db	1201	GACAAAGTGAAGACCGCGCTCATTCGAGACCGCGGACATATCGCCCGCGAAGGATAGCG	1260
Qy	1261	GACATCGCTACGTTGGCGGCTAGGGTGAACGCTCTCAAGGCGCATCAAGTACGACGACTAC	1320
Db	1261	GACATCGCTACGTTGGCGGCTAGGGTGAACGCTCTCAAGGCGCATCAAGTACGACGACTAC	1320
Qy	1321	GCCAAAGTCAACGCTTCCGCGCGGACAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG	1380
Db	1321	GCCAAAGTCAACGCTTCCGCGCGGACAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG	1380
Qy	1381	GTGAGCG	1440
Db	1381	GTGAGCG	1440
Qy	1441	GACCTTACCTTACGACCG	1500
Db	1441	GACCTTACCTTACGACCG	1500
Qy	1501	GGCTTCGAGAGGTCGGGTACTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1560
Db	1501	GGCTTCGAGAGGTCGGGTACTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1560
Qy	1561	AGCTACAAGCG	1620
Db	1561	AGCTACAAGCG	1620
Qy	1621	TCCG	1680
Db	1621	TCCG	1680
Qy	1681	CAGACCTTACCGCGGTTCGTTAAGGACTTACGAGACCGCGCGCGCGCGCGCGCGCGCG	1740
Db	1681	CAGACCTTACCGCGGTTCGTTAAGGACTTACGAGACCGCGCGCGCGCGCGCGCGCGCG	1740
Qy	1741	GTCAACAGGCG	1800
Db	1741	GTCAACAGGCG	1800
Qy	1801	CTGACGCTTACCTTACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1860
Db	1801	CTGACGCTTACCTTACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1860
Qy	1861	AGCTACGAGCAGCTGCGAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1920
Db	1861	AGCTACGAGCAGCTGCGAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1920
Qy	1921	GCCTACAGCAGCTTACGCGGTGGCGGACTACAGCTCAAGCGCGCGCGCGCGCGCGCG	1977
Db	1921	GCCTACAGCAGCTTACGCGGTGGCGGACTACAGCTCAAGCGCGCGCGCGCGCGCGCG	1977

Qy	1	ATGAAGGGGCTGAAAGCTCTCATATTAGTCAATTTAGTTCAGGTTTGGTAGTAGGAGC	60
Db	1	ATGAAGGGGCTGAAAGCTCTCATATTAGTCAATTTAGTTCAGGTTTGGTAGTAGGAGC	60
Qy	61	GTAGCGGAGCTCCAGAGCAAGATTGAACAAAGTAAGAAATTTGTAAGAAAGTAATGCT	120
Db	61	GTAGCGGAGCTCCAGAGCAAGATTGAACAAAGTAAGAAATTTGTAAGAAAGTAATGCT	120
Qy	121	CTGTAACGCGAGGACTGTTCAGAAAATTTCAAAAATTTCAATTCCTAACGAGGAATCAGC	180
Db	121	CTGTAACGCGAGGACTGTTCAGAAAATTTCAAAAATTTCAATTCCTAACGAGGAATCAGC	180
Qy	181	ACAGTAATTTGTAATTTGAAAGCAATAGGGAAGAAATTTGCAAGTAAGAGTTCTTGAGTA	240
Db	181	ACAGTAATTTGTAATTTGAAAGCAATAGGGAAGAAATTTGCAAGTAAGAGTTCTTGAGTA	240
Qy	241	ATGGGTGCAAAAGTTAGTGTGCTGACCATATTATACCGCAATAGCTGCGGATCTTAAG	300
Db	241	ATGGGTGCAAAAGTTAGTGTGCTGACCATATTATACCGCAATAGCTGCGGATCTTAAG	300
Qy	301	GTGAGAGCTTACTAGTCAATCTCAGGTTTACAGGGGGTAAGCTTAAGCTTTCAGGTTT	360
Db	301	GTGAGAGCTTACTAGTCAATCTCAGGTTTACAGGGGGTAAGCTTAAGCTTTCAGGTTT	360
Qy	361	AGGTTTATCAGAGCAAGCTACAAAGTTACAGTTTTCAGAGCAATTAAGAGGACTGGATGA	420
Db	361	AGGTTTATCAGAGCAAGCTACAAAGTTACAGTTTTCAGAGCAATTAAGAGGACTGGATGA	420
Qy	421	TCTGAGCTCAAGTTATGCAACTTACCTTTGGAGCTTGGATGATGATGTTTCTGGAATC	480
Db	421	TCTGAGCTCAAGTTATGCAACTTACCTTTGGAGCTTGGATGATGATGTTTCTGGAATC	480
Qy	481	ACAATAGGAATTAATGACACTTGAATTTGACGCTTCTCATCCAGATCTCAGAGAAAGTA	540
Db	481	ACAATAGGAATTAATGACACTTGAATTTGACGCTTCTCATCCAGATCTCAGAGAAAGTA	540
Qy	541	ATGGGTGGGTAGATTTGTCAATGTTAGGAGTTATCCATACGATGACCATGGACATGGA	600
Db	541	ATGGGTGGGTAGATTTGTCAATGTTAGGAGTTATCCATACGATGACCATGGACATGGA	600
Qy	601	ACTCATGAGCTTCAATAGCAGTGGTACTGGAGCAGCAAGTAATGGCAGTACAGGGA	660
Db	601	ACTCATGAGCTTCAATAGCAGTGGTACTGGAGCAGCAAGTAATGGCAGTACAGGGA	660
Qy	661	ATGGCTCCAGGAGCTAGCTTGGCGGAAATTAAGGTTCTAGGTGCGGATGTTTCTGGAAGC	720
Db	661	ATGGCTCCAGGAGCTAGCTTGGCGGAAATTAAGGTTCTAGGTGCGGATGTTTCTGGAAGC	720
Qy	721	ATATCTACTAATTAAGGAGTTGATGCGCGCTTGATTAACAAAGATAAGTACGGAAT	780
Db	721	ATATCTACTAATTAAGGAGTTGATGCGCGCTTGATTAACAAAGATAAGTACGGAAT	780
Qy	781	AAAGCTAATTAATTTCTTCTTGGTTCAAGCCAGAGCTCCGAGCGGACGACTCCCTCAGT	840
Db	781	AAAGCTAATTAATTTCTTCTTGGTTCAAGCCAGAGCTCCGAGCGGACGACTCCCTCAGT	840
Qy	841	CAGGCGCTCAACACGCTGGGACCGCGGTATAGTAGTCTGCGTGGCGCGCGCAACAGC	900
Db	841	CAGGCGCTCAACACGCTGGGACCGCGGTATAGTAGTCTGCGTGGCGCGCGCAACAGC	900

RESULT 2

AAT85695

ID AAT85695 standard; DNA; 1962 BP.

✕✕

AC AAT85695;

XXXXXX

DT 20-APR-19

XX
XX

XX
DE
FY1000000:

KW Protease:

xx
NY
Process;

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Pyrococcus

6
0
C
7
E
C
0
:

X

PN WO9721823

~~_____~~
~~_____~~

19-JUN-1961

1

3d
67 - NOV - 1967

XX

PR 12-DEC-199

7			
4			
2			
1			
	X	X	X

PA (TAKI, TA

XX
FH (TWT)
TWT (TWT)

DT
YY

PI ASADA K,
PI TULLOCH

PI TSUNASAWA

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

WPI: 1997

DR P-PSDB; AA

XX

PT Procure(s)

PT **Pyrococcus**

PT **useful** **inc**

XX

PS Disclosure

XXXXXX

CC This sequence

Pyrococcus

CC
CC
CC

stability

Industrial
Stability:

**XX
CC
INDUSTRIAL**

QY	361	AGGTTTATCCAGGAGACTACAAGTTACAGTTTACAGACAGAAATTAGAAGGACTGGATGAG	420
Db	361	AGGTTTATCCAGGAGACTACAAGTTACAGTTTACAGACAGAAATTAGAAGGACTGGATGAG	420
QY	421	TCTCCAGCTCAAGTTATGGCAACTTACGTTTGGAACTTGGGATATGATGGTTCTCGGAATC	480
Db	421	TCTCCAGCTCAAGTTATGGCAACTTACGTTTGGAACTTGGGATATGATGGTTCTCGGAATC	480
QY	481	ACAATAGGAATAATTGACACTGGAAATTGACGCTTCTCATCCAGATCTCCAAAGCAAAAGTA	540
Db	481	ACAATAGGAATAATTGACACTGGAAATTGACGCTTCTCATCCAGATCTCCAAAGCAAAAGTA	540
QY	541	ATTGGGTGGGTAGATTTTGTCATAGTGTAGGAGTTTTCATACGATGACCATGACATGGA	600
Db	541	ATTGGGTGGGTAGATTTTGTCATAGTGTAGGAGTTTTCATACGATGACCATGACATGGA	600
QY	601	ACTCATGTAGCTTCAANTAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAAAGGA	660
Db	601	ACTCATGTAGCTTCAANTAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAAAGGA	660
QY	661	ATGGCTCCAGAGACTAAGCTTGGCGGGAATTAAGGTTCTAGTGGCGATGGTCTCGAAGC	720
Db	661	ATGGCTCCAGAGACTAAGCTTGGCGGGAATTAAGGTTCTAGTGGCGATGGTCTCGAAGC	720
QY	721	ATATCTACTATTAATTAGGAGTTGAGTGGGCGTTGTATACAAAGATAAGTACGGAATT	780
Db	721	ATATCTACTATTAATTAGGAGTTGAGTGGGCGTTGTATACAAAGATAAGTACGGAATT	780
QY	781	AAGTCAATTAACTTCTCTTGGTTCAAGCCAGAGCTCCGAGCGAAGCACTCCCTCAGT	840
Db	781	AAGTCAATTAACTTCTCTTGGTTCAAGCCAGAGCTCAGATGTTACTGAGCCTCTAAGT	840
QY	841	CAGGCGCTCAACAACGCTGGGACGCGGTATAGTACTGGTCCGCGCGGCAACAGC	900
Db	841	CAGGCTTAAATGCAGCGTGGGATCGTGATTAAGTTTGTGTGTGTCGCGTGGAAACACT	900
QY	901	GGGCGCAAGACCTTACACGCTCGGCTCACCGCGCGCGGCGGAGCAAGTCAATACCGTCGGT	960
Db	901	GGACCTTACACAGTATACATTCGGTTCTCCAGAGCTGCAAGCAAGTTATTACAGTTGGA	960
QY	961	GCAGTTGACGCAACGACACACTCGCCAGCTTCTCAGCAGGAGCAGCGCGGACACGA	1020
Db	961	GCGGTTGACCAAGTATGATGTTTAAACAAGTCTCTAAGCAGAGGGCCAACTGCACAGCGC	1020
QY	1021	AGGCTCAAGCGGGAAGTCGTCGCCCGGGGTTGACATCATAGCCGCGGCGCAGCGGA	1080
Db	1021	AGGCTTAAGCCTAGGTTGTTGCTCCAGGAACCTGGATTAATGCTGCGCAGCAAGTGGGA	1080
QY	1081	ACCAGATGGGACCCCGATAAACGACTACTACACAAGGCTCTGGAACCAACGATGCCC	1140
Db	1081	ACTAGCTGGGTCAACCAATTAACTACTATTATACAGAGCTCTCGGACATCAATAGCA	1140
QY	1141	ACCCCGACAGTTTCGGGCGTTGGCGGCTCATCTCTCAGAGGCCCAACCGAGCTGGACGCGG	1200
Db	1141	ACTCCTCAGTAGCTGATATTCAGGCGCTCTTGCTCCAAGSCACACCGAGCTGGACTCCA	1200
QY	1201	GACAAGGTGAAGACCGCCCTCATGAGACCGCGCATAGTTCGCCGCCCAAGGAGATACGG	1260
Db	1201	GACAAAGTAAACACGCCCTCATAGAARCTGCTGTATCTGTAAGAGCCAGATGAATAGCC	1260
QY	1261	GACATCGCCTACGGTGGGGTAGGTTGAACGCTCTACAAGGCCATCAAGTACGACGACTAC	1320
Db	1261	CATATAGGCTTACGGTGCAGTAGGTTTAATGCAATACAAGGCTATAAATACGATAACTAT	1320
QY	1321	GCCAAAGCTCACTTCAACCGGCTCGGTCGGCAGCAAGGAGCGCACCCACACCTTCGCAC	1380
Db	1321	GCAAGCTAGTGTTCACTGATATGTTTGCCACAAAGCGACCAACTCACCAGTTGTTT	1380
QY	1381	GTACGCGCGGCCACTTTCGTGACCGCACCCACCTCTACTGGGACAGGGGCTCGAGCGACATC	1440
Db	1381	ATTAGCGGAGGTTCTGTTGACTGTCACCAATATATCTGGACAACTGCCAATAAGGACCTT	1440

CC The invention relates to a hyperthermostable protease derived from a thermophilic bacterium (especially Pyrococcus furiosus). The protease has working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10 (optimum 6-8), and retains more than 90% of its activity after 8 hours at 95 deg.C. The invention also provides gene sequences encoding a polypeptide of formula STG-Ala-Gly-Asn-PRO, where STG is a signal peptide from subtilisin, and PRO is the above protease. Host cells (especially Bacillus strains) transformed with vectors comprising the genes are used for the recombinant production of the protease. The hyperthermostable protease which can be prepared in quantity suitable for industrial use, can be used as an additive for drugs, washing agents and foodstuffs and for chemical synthesis.

Query Match 69.1%; Score 1366.4; DB 20; Length 1962;
Best Local Similarity 81.2%; Pred. No. 2.8e-283;
Matches 1605; Conservative 0; Mismatches 356; Indels 15; Gaps 1;

QY 1441 GACCTTACTCTAGACCCCAACGAGGTTGACTACTCTTACACCGCTACTAC 1500
DB 1441 GATCTTTACTCTAGCTCCCAATGGAACCGAGTTGACTCTTACACCGCTACTAT 1500
QY 1501 GGCTTCGAGAAGTGGGTACTACAAACCGACCGCCGGAACCTCGAGCGGTCAAGGTGTC 1560
DB 1501 GGATTCGAGAAGTGGGTATTAACACCCCACTGATGGAACATGGACAATTAAGGTGTA 1560
QY 1561 AGCTACAGGGCGCGCACTACAGGTGCGTGCAGCGAGCGGAGCCCTCACGCCAG 1620
DB 1561 AGCTACAGGGCGAGTGCACATATCAAGTAGATGCTGTAAGTGTGCTTCCCTTCACAG 1620
QY 1621 TCCGGCGGGCGCAACCGCAATCCAAACCCCAACCGCAACCCCGACCGACCGAC 1680
DB 1621 CCTGGAAG-----TTCACCATCTCCCAACCCCAACCGAGTAGACGCA 1665
QY 1681 CAGACCTTCACCGGTTCCGTTAAGGACTAGTGGACACCGACGACACCTTCACCATGAAC 1740
DB 1681 AAGACGTTCCAGGATCCGATCCTACTACTATGACAGGAGCGACACCTTCAATGACC 1725
QY 1741 GTCACACGCGTCCACCAAGATACCGGTGACCTCACCTGCGATACTCTTCAACACGAC 1800
DB 1741 GTCACACGCGTCCACCAAGATACCGGTGACCTCACCTGCGATACTCTTCAACACGAC 1800
QY 1726 GTTAACCTCTGGGCTACAAAGATACCTGAGACCTAGTGTGACACGAGCTACCATGAT 1785
DB 1726 GTTAACCTCTGGGCTACAAAGATACCTGAGACCTAGTGTGACACGAGCTACCATGAT 1785
QY 1801 CTCACCTCTACCTCTACGACCCCAACGCGCACTCTGTTGACAGGTCCACGTCCAGCAAC 1860
DB 1801 CTCACCTCTACCTCTACGACCCCAACGCGCACTCTGTTGACAGGTCCACGTCCAGCAAC 1860
QY 1786 CTTGACCTTTACCTCTACGATCTTACCAAGGCTTGTATAGATAGATCGGAGAGTCCCAAC 1845
DB 1786 CTTGACCTTTACCTCTACGATCTTACCAAGGCTTGTATAGATAGATCGGAGAGTCCCAAC 1845
QY 1861 AGCTACGAGCAGCTCCAGTACGCAACCCCGGCAACCTGGACCTTCCTGCTCTAC 1920
DB 1861 AGCTACGAGCAGCTCCAGTACGCAACCCCGGCAACCTGGACCTTCCTGCTCTAC 1920
QY 1846 AGCTACGAGCAGCTACGATATTAACCTCCGCGCCAGGACCTGCTACTCTCTAGTATAT 1905
DB 1846 AGCTACGAGCAGCTACGATATTAACCTCCGCGCCAGGACCTGCTACTCTCTAGTATAT 1905
QY 1921 GCTACGACGACCTACGCTGGCGGAGTACGAGTCAAGGCGCTGCTCTACTAGG 1976
DB 1906 GCTACTACGCTTACGCTTGGCTTACTACGAGCTGACGCTAAAGTTTATTATGG 1961

RESULT 3
AA05929
ID AA05929 standard; DNA; 1962 BP.
XX
AC AAX05929;
XX
XX 06-MAY-1999 (first entry)
XX
XX Hyperthermostable protease encoding DNA.
XX
XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
XX additive; drug; washing agent; foodstuff; chemical synthesis; ds.
XX
XX Pyrococcus furiosus.
XX
XX WO9856926-A1.
XX
XX 17-DEC-1998.
XX
XX 04-JUN-1998; 98WO-JP02465.
XX
XX 10-JUN-1997; 97JP-0151969.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX Asada K, Kato I, Morishita M, Shimojo T, Takakura H;
XX
XX WPI; 1999-080907/07.
XX P-PSDB; AAW94841.
XX
XX Recombinant hyperthermostable protease from Pyrococcus furiosus.
XX and gene encoding it, for large scale production of the protease for
XX industrial use.
XX
XX Disclosure; Page 59-60; 82pp; Japanese.
XX
XX

CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10 (optimum 6-8), and retains more than 90% of its activity after 8 hours at 95 deg.C. The invention also provides gene sequences encoding a polypeptide of formula STG-Ala-Gly-Asn-PRO, where STG is a signal peptide from subtilisin, and PRO is the above protease. Host cells (especially Bacillus strains) transformed with vectors comprising the genes are used for the recombinant production of the protease. The hyperthermostable protease which can be prepared in quantity suitable for industrial use, can be used as an additive for drugs, washing agents and foodstuffs and for chemical synthesis.

Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 other;

Query Match 69.1%; Score 1366.4; DB 20; Length 1962;
Best Local Similarity 81.2%; Pred. No. 2.8e-283;
Matches 1605; Conservative 0; Mismatches 356; Indels 15; Gaps 1;

QY 1 AUAAGGGGCTGAAGCTCTCATATTAGTGATTTTAGTTCTTAGTTTGGTAGGGAGC 60
DB 1 AUAAGGGGCTGAAGCTCTCATATTAGTGATTTTAGTTCTTAGTTTGGTAGGGAGC 60
QY 61 GTAGCGGAGCTCCAGAGAAAGTGAACAAAGTAAGTAAGTAAGTGTGAGAAAGTATGT 120
DB 61 GTAGCGGAGCTCCAGAGAAAGTGAACAAAGTAAGTAAGTAAGTGTGAGAAAGTATGT 120
QY 121 CTGCTAACCGCAGGAGCTCTCAGAAATTCAGAAATTCAGAAATTCAGAAATTCAGC 180
DB 121 CTGCTAACCGCAGGAGCTCTCAGAAATTCAGAAATTCAGAAATTCAGAAATTCAGC 180
QY 181 ACAGTAATTTGATTTGAAACCATAGGAAAGTAAGTAAGTGTGAGTAAGTCTTGAATTA 240
DB 181 ACAGTAATTTGATTTGAAACCATAGGAAAGTAAGTAAGTGTGAGTAAGTCTTGAATTA 240
QY 241 ATGGGTGCAAAAGTATGCTATGCTACCATATATATACCGCAATAGTGTCCGATCTTAAG 300
DB 241 ATGGGTGCAAAAGTATGCTATGCTACCATATATATACCGCAATAGTGTCCGATCTTAAG 300
QY 301 GTTAGAGCTTACTAGTCTCAGGTTTACAGGGGTAAAGCTTAAGCTTACGCTTTCAGGTGT 360
DB 301 GTTAGAGCTTACTAGTCTCAGGTTTACAGGGGTAAAGCTTAAGCTTACGCTTTCAGGTGT 360
QY 361 AGGTTTATCCAGGAGACTACAAAGTTACAGTTTACAGCAATTAGAGAGCTGGATGAG 420
DB 361 AGGTTTATCCAGGAGACTACAAAGTTACAGTTTACAGCAATTAGAGAGCTGGATGAG 420
QY 421 TCTCCAGCTCAGGTTATGCAACTTACGTTTGGAACTTGGGATATGATGTTCTGGAATC 480
DB 421 TCTCCAGCTCAGGTTATGCAACTTACGTTTGGAACTTGGGATATGATGTTCTGGAATC 480
QY 481 ACAATAGGAATAATTGACACTGGAAATTCAGCTTCTCATCCAGATCTCCAGGAAGTA 540
DB 481 ACAATAGGAATAATTGACACTGGAAATTCAGCTTCTCATCCAGATCTCCAGGAAGTA 540
QY 541 ATTGGGTGGGTAGATTTGTCATGCTAGGAGTATTCATACGATGACCATGGAATGGA 600
DB 541 ATTGGGTGGGTAGATTTGTCATGCTAGGAGTATTCATACGATGACCATGGAATGGA 600
QY 601 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAAGTAAGTAAAGTAAAGTAAAGTAA 660
DB 601 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAAGTAAGTAAAGTAAAGTAAAGTAA 660
QY 661 ATGGCTCCAGGAGCTAAAGCTGGCGGAATTAAGGTTCTAGGTGCCGATGGTCTCGAAGC 720
DB 661 ATGGCTCCAGGAGCTAAAGCTGGCGGAATTAAGGTTCTAGGTGCCGATGGTCTCGAAGC 720
QY 721 ATATCTACTATATTAAGGAGTGTGAGTGGCCGCTTCATTAACAAGATAAGTACGGAAT 780
DB 721 ATATCTACTATATTAAGGAGTGTGAGTGGCCGCTTCATTAACAAGATAAGTACGGAAT 780
QY 781 AAGGTCATTAATCTTCTCTTGGTTCACAGGCTCCGACGAGCTCCGACGAGCTCCCTCAGT 840
DB 781 AAGGTCATTAATCTTCTCTTGGTTCACAGGCTCCGACGAGCTCCGACGAGCTCCCTCAGT 840

QY	1484	CCTTACACGGCTACTACGGCTTCGAGAAAGTTCGGTCTACTACAAACCGACCGCCGGAAACCT	1543
Db	1484	CCTTACACGGCTACTACGGCTTCGAGAAAGTTCGGTCTACTACAAACCGACCGCCGGAAACCT	1543
QY	1544	GGACGGTCAAGGTGCTGACTACAAAGGGCGCGCACTACCAAGTTCGACGCTCGTCAGCG	1603
Db	1544	GGACGGTCAAGGTGCTGACTACAAAGGGCGCGCACTACCAAGTTCGACGCTCGTCAGCG	1603
QY	1604	ACGGGAGGCTCAGCCAGTCCGGCGGGGCAACCGCAATCCAAACCCCAACCGCAACCCAA	1663
Db	1604	ACGGGAGGCTCAGCCAGTCCGGCGGGGCAACCGCAATCCAAACCCCAACCGCAACCCAA	1663
QY	1664	CCCCGACCAACCGACACCCAGACTTTCACCGGTTTCGGTTAAGCACTACTCGGACACCAAGCG	1723
Db	1664	CCCCGACCAACCGACACCCAGACTTTCACCGGTTTCGGTTAAGCACTACTCGGACACCAAGCG	1723
QY	1724	ACACCTTCACCTGAACGCTCAACACGGGTGCACCAAGATAACCGGTGACCTTGACCTTCG	1783
Db	1724	ACACCTTCACCTGAACGCTCAACACGGGTGCACCAAGATAACCGGTGACCTTGACCTTCG	1783
QY	1784	ATACTTCTCAACAGCACTCGACCTCTACCTCTACGACCCCAACGGCAACCTTCGTTGACA	1843
Db	1784	ATACTTCTCAACAGCACTCGACCTCTACCTCTACGACCCCAACGGCAACCTTCGTTGACA	1843
QY	1844	GTCCACGTCGAGCAACAGCTACGACGCTCGAGTAGTACGCCAAACCCGCCCGGGGAACCT	1903
Db	1844	GTCCACGTCGAGCAACAGCTACGACGCTCGAGTAGTACGCCAAACCCGCCCGGGGAACCT	1903
QY	1904	GGAGGTCCTCGTCTACGCTTACGACACCTACGACACCTACGGCTGGCGGACTTACACGCTCAGGCGG	1963
Db	1904	GGAGGTCCTCGTCTACGCTTACGACACCTACGACACCTACGGCTGGCGGACTTACACGCTCAGGCGG	1963
QY	1964	TCGTCCTACTACGGG	1977
Db	1964	TCGTCCTACTACGGG	1977
RESULT 5			
AAK05926			
ID	AAK05926 standard; DNA: 1977 BP.		
XX	AC	AAK05925;	
XX	XX		
XX	XX	06-MAY-1999 (first entry)	
XX	DE	W09856926 Seq ID 11.	
XX	KW	Hyperthermostable; protease; thermophilic; bacterium; subtilisin;	
XX	KW	additive; drug; washing agent; foodstuff; chemical synthesis; ds.	
XX	OS	Thermococcus caler.	
XX	PN	W09856926-A1.	
XX	PD	17-DEC-1998.	
XX	PF	04-JUN-1998; 98W0-JP02465.	
XX	PR	10-JUN-1997; 97JP-0151969.	
XX	PA	(TAKI) TAKARA SHUZO CO LTD.	
XX	PI	Asada K, Kato I, Morishita M, Shimojo T, Takakura H;	
XX	DR	WPI: 1999-080907/07.	
XX	DR	P-PSDB; AAN94840.	
XX	PT	Recombinant hyperthermostable protease from <i>Pyrococcus furiosus</i> -	
XX	PT	and gene encoding it, for large scale production of the protease for	
XX	PT	industrial use.	
XX	PS	Disclosure; Page 53-54; 82pp; Japanese.	
XX	XX		

[illegible]

The invention relates to a hyperthermostable protease derived from a thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10 (optimum 6-8), and retains more than 90% of its activity after 8 hours at 95 deg.C. The invention also provides gene sequences encoding a polypeptide of formula $\text{SIG-ALA-Gly-Asn-PRO}$, where SIG is a signal peptide from subtilisin, and PRO is the above protease. Host cells (especially *Bacillus* strains) transformed with vectors comprising the genes are used for the recombinant production of the protease. The hyperthermostable protease which can be prepared in quantity suitable for industrial use, can be used as an additive for drugs, washing agents and foodstuffs and for chemical synthesis.

Query Match	68.8%	Score 1360.8	DB 20	Length 1977
Best Local Similarity	83.9%	Pred. NO. 4,5e-282		
Matches 1573	Conservative 0	Mismatches 292	Indels 9	Gaps 3

Qy	107	AGAAGAC	ATATGGTCTGCTAACGCCAGGAGCTGTTTCAGAAAAATTCAAAAAATTTGAATCCTTA	166
Db	110	AGAAGAC	ATACGGACTCTGTACCCCGGAGCTGTTCAAGAAAATTTCCAGAGAGTGAACCTGGGA	169
Qy	167	ACGAGAAAT	CAGACACAGTAATTTGTATTTGAAAACCACTACGGGAAGAAAGAAATTTCCAGTAA	226
Db	170	ACCAGAACT	GGACACCGTCAATATGTTCTGGGAGCTACGGAGACAGGACAGGCGGCTTTA	229
Qy	227	GAGTCTTCACTT	AATGGGTGCMAAGTTTAGTATGTGTACCATATTTATACCCCAATAG	286
Db	230	AGTACTGAGGCT	CATGGGGCCCGAGGTCAAGTACTCTACAGATATATCCCTGCTGTCGTCG	289
Qy	287	CTCCGATCTT	TAAGGTTTAGAGACTTTACTAGTCACTCTCAGGTTTAACAGGGGTAAAGCTA	346
Db	290	CGGTTAAAT	TAAGCCAGGAGCCCTCTGCTGATCGGGGCGCATGATA---GACACGGGTT	346
Qy	347	AGCTTTCAGCT	TTTAGTTTATCCAGGAAGACTACAAAGTTACAGTTTCACGAGAAATTAG	406
Db	347	ACTTCGGT	TACACAAAGGCTCTGGGCGATAAAGTTTCATACAGGAGGATTTACAAAGTTTCAGG	406
Qy	407	AAGGACTCGAT	GTGAGTCTGCGAGCTCAAGTTATGCGAACTTACGTTTGGAACT---TGGGAT	463
Db	407	TTGACGAGCCACT	TTCCGCTCTCCAGATAGGGGCGGATACCGTCTGGAACCTCCCTCGGCT	466
Qy	464	ATCATGTTCT	TGGAATCAACAATAGGAATAATTGACACTGGAATTGAGCGCTTCTCATCCAG	523
Db	467	ACGACGGAAGG	GGTGGTGGCTGCTGCTCGATACCGGTTATAGCGGATAGACGGCAACCCCG	526
Qy	524	ATCTCCAAGGA	AAAGTAATTTGGTGGGTAGATTTTCTCAATGGTAGGAGTTATCCATACG	583
Db	527	ATCTGAGGGC	AGGTCTATAGGCTGGTACGACGCCCTCAAGGCGAGGTCGACGCCCTACG	586
Qy	584	ATGACCATG	GACATGGAACCTCATGTAGCTTCAATACAGCTGGTACTGGAGCAGCAAGTA	643
Db	587	ATGACCAAGG	ACAGCAACCCACGTTGCGGTTATCTTCCGGAAACCGGCAGCG---TTA	643
Qy	644	ATGGCAAGT	ACAAGGGAATGGCTCAGGAGCTAAGCTGCGGGGAATTTAAGGTTCTTAGTG	703
Db	644	ACTCCAGT	ATACATAGCGTTCGCCCCCGCGCGAAGCTCTCGGGCTCAAGGTTCTCGGT	703
Qy	704	CCCATGGT	CTCTGGAAGCATATCTACTATATTTAAGGAGCTTAGTGGGCGGTTGATAACA	763
Db	704	CCGACGGT	TCGGGAAGCTCTCCACATCATCGCGGGGTGTGACTGGGTGTCGAGAACA	763
Qy	764	AGATATAGT	ACGGAAATTAAGTGCTAATTTCTCTTTGGTTCAAGCCAGAGCTCCGACG	823
Db	764	AGGACAAGT	ACGGGAATTAAGGTCATCAACCTCTCCCTCGGCTCTCCCTCCAGAGCTCCGACG	823
Qy	824	GAACCGACT	CCCTCTACTCAGGCCGCTCAACAAGCGCTGGAGCGCGGTTAGTAGTCTGCG	883
Db	824	GAACCGACT	CCCTCTACTCAGGCCGCTCAACAAGCGCTGGAGCGCGGTTAGTAGTCTGCG	883
Qy	884	TCGCGCGCG	GCACACCGCGCGGCGGCAACCTACACGCTCGGCTCACCCGCCCGCGGAGCA	943

Thu Nov 7 10:12:16 2002

RESULT 5
AAT85677
ID AAT85677 standard; DNA; 1859 BP.
XX AC AAT85677;
XX DT 20-APR-1998 (first entry)
XX DE Thermococcus protease fragment coding sequence.
XX KW Protease; research reagent; thermal stability; thermococcus celer; ss.
XX OS Thermococcus celer DSM-2476.
XX PN WO9721823-A1.
XX PD 19-JUN-1997.
XX PF 07-NOV-1996; 96WO-020253.
XX PR 12-DEC-1995; 95JP-0323285.
XX PA (TAKI) TAKARA SHUZO CO LTD.
XX PI Asada K, Kato I, Mita M, Morishita M, Takakura H;
XX PL Tsunashima S, Yamamoto K;
XX DR WPI; 1997-332794/30.
XX PT Protease(s) and genes encoding them obtained from Thermococcus and
XX PT Pyrococcus strains - have extremely high thermal stability and are
XX PT useful industrially and as research reagents
XX PS Disclosure; Page 112-113; 159pp; Japanese.
XX CC This sequence represents a fragment of the coding sequence for the
XX CC protease from Thermococcus celer DSM-2476 (see AAT85677 for full length
XX CC sequence). This sequence encodes a fragment of the protease of the
XX CC invention. The proteases of the invention have extremely high thermal
XX CC stability. The proteases can be used as research reagents, and
XX CC industrially in the food, drug and chemical industries.
XX SQ Sequence 1859 BP; 420 A; 631 C; 509 G; 299 T; 0 other;
Query Match 58.9%; Score 1165; DB 18; Length 1859;
Best Local Similarity 100.0%; Pred. No. 4.1e-240;
Matches 1165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 813 GAGCTCCGACGGACCGACCTCCCTCAGTCAGGCGGTCAACAGCGCTGGAGCGCGGTAT 872
DB 1 GAGCTCCGACGGACCGACCTCCCTCAGTCAGGCGGTCAACAGCGCTGGAGCGCGGTAT 60
QY 873 AGTAGTCTCGCTCG 932
DB 61 AGTAGTCTCGCTCG 120
QY 933 CGGT 992
DB 121 CGGT 180
QY 993 CTCAGAGGCGGT 1052
DB 181 CTCAGAGGCGGT 240
QY 1053 TGACATCATAGCGT 1112
DB 241 TGACATCATAGCGT 300
QY 1113 CACCAGGCGCTGTGAACACGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 1172
DB 301 CACCAGGCGCTGTGAACACGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 360

QY 1173 CCTCAGGCG 1232
DB 361 CCTCAGGCG 420
QY 1233 CGACATAGTCTCG 1292
DB 421 CGACATAGTCTCG 480
QY 1293 CTACAAGGCG 1352
DB 481 CTACAAGGCG 540
QY 1353 CAAAGGCG 1412
DB 541 CAAAGGCG 600
QY 1413 CTACTGGGACACGGGCTCGAGCGGATCGAGCGGATCGAGCGGATCGAGCGGATCG 1472
DB 601 CTACTGGGACACGGGCTCGAGCGGATCGAGCGGATCGAGCGGATCGAGCGGATCG 660
QY 1473 GGTGACTTACTCTTACACCGCGCTTACCGCGCTTACCGCGCTTACCGCGCTTACCG 1532
DB 661 GGTGACTTACTCTTACACCGCGCTTACCGCGCTTACCGCGCTTACCGCGCTTACCG 720
QY 1533 CGCGGGAACCTGGAGCGGTCAAGGTGCTCAGCTACAGGGCGCGCGCGCGCGCGCG 1592
DB 721 CGCGGGAACCTGGAGCGGTCAAGGTGCTCAGCTACAGGGCGCGCGCGCGCGCGCG 780
QY 1593 CGTCTCAGCGGACGGGAGCTCAGCGGAGCTCAGCGGAGCTCAGCGGAGCTCAGCG 1652
DB 781 CGTCTCAGCGGACGGGAGCTCAGCGGAGCTCAGCGGAGCTCAGCGGAGCTCAGCG 840
QY 1653 CCGGGAACCG 1712
DB 841 CCGGGAACCG 900
QY 1713 GGACCGGCG 1772
DB 901 GGACCGGCG 960
QY 1773 CTTGACCTTGGATCTTCTTACAGCGCTCAGCGGAGCTCAGCGGAGCTCAGCGG 1832
DB 961 CTTGACCTTGGATCTTCTTACAGCGCTCAGCGGAGCTCAGCGGAGCTCAGCGGAG 1020
QY 1833 CCTCGTTCAGGCTTCAGCGTTCAGCGTTCAGCGTTCAGCGTTCAGCGTTCAGCGTTC 1892
DB 1021 CCTCGTTCAGGCTTCAGCGTTCAGCGTTCAGCGTTCAGCGTTCAGCGTTCAGCG 1080
QY 1893 CCGCGGAACCTGGAGCGGTTCCTGCTTACCGCTTACAGCGCTTACAGCGCTTACAG 1952
DB 1081 CCGCGGAACCTGGAGCGGTTCCTGCTTACCGCTTACAGCGCTTACAGCGCTTACAG 1020
QY 1953 GCTCAAGGCGCGTGGTCTACTACCGG 1977
DB 1141 GCTCAAGGCGCGTGGTCTACTACCGG 1165

RESULT 7
AAT85668
ID AAT85668 standard; DNA; 1566 BP.
XX AC AAT85668;
XX DT 20-APR-1998 (first entry)
XX DE Pyrococcus furiosus protease coding sequence.
XX KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX OS Pyrococcus furiosus DSM-3638.
XX FH Key Location/Qualifiers

AC AAT08131;
 DT 10-MAY-1996 (first entry)
 DE Hyperthermostable protease gene.
 XX Protease; hyperthermostable; thermostability; ss.
 OS Pyrococcus furiosus.
 FH Key Location/Qualifiers
 FT misc_difference 587
 FT /*tag= a
 FT /*note= "base n at position 587 is not identified
 FT in the specification"
 FT misc_difference 718
 FT /*tag= b
 FT /*note= "base n at position 718 is not identified
 FT in the specification"
 FT misc_difference 784
 FT /*tag= c
 FT /*note= "base n at position 784 is not identified
 FT in the specification"
 FT misc_difference 796
 FT /*tag= d
 FT /*note= "base n at position 796 is not identified
 FT in the specification"
 FT misc_difference 801
 FT /*tag= e
 FT /*note= "base n at position 801 is not identified
 FT in the specification"
 FT misc_difference 815
 FT /*tag= f
 FT /*note= "base n at position 815 is not identified
 FT in the specification"
 FT misc_difference 859
 FT /*tag= g
 FT /*note= "base n at position 859 is not identified
 FT in the specification"
 FT misc_difference 878
 FT /*tag= h
 FT /*note= "base n at position 878 is not identified
 FT in the specification"
 FT misc_difference 887
 FT /*tag= i
 FT /*note= "base n at position 887 is not identified
 FT in the specification"
 FT misc_difference 890
 FT /*tag= j
 FT /*note= "base n at position 890 is not identified
 FT in the specification"
 FT misc_difference 896
 FT /*tag= k
 FT /*note= "base n at position 896 is not identified
 FT in the specification"

W09534645-A1
 21-DEC-1995.

05-JUN-1995: 95W0-JP01095.
 26-JUL-1994: 94JP-0173912.
 13-JUN-1994: 94JP-0130236.
 (TAKI) TAKARA SHUZO CO LTD.
 Asada K, Kato I, Mita M, Morishita M, Tsunawasa S;
 Yamamoto K;
 WPI: 1996-049674/05.
 P-PSDB; AAR87009.

PT Pyrococcus furiosus hyperthermostable protease gene - useful for
 XX recombinant prodn. of hyperthermostable protease
 PS Example 3: Page 52-53; 85pp; Japanese.
 XX The invention relates to the hyperthermostable protease of
 CC Pyrococcus furiosus and its prodn. as a recombinant protein
 CC in transformants using a vector carrying the protease gene
 CC (AAR08141). A genomic DNA sequence of the invention is given in
 CC AAT08131 and its encoded protein in AAR87009.
 XX
 SQ Sequence 898 BP; 192 A; 299 C; 266 G; 130 T; 11 other;
 Query Match 33 9%; Score 669.8; DB 17; Length 898;
 Best Local Similarity 86.3%; Pred. No. 4.7e-134;
 Matches 780; Conservative 0; Mismatches 118; Indels 6; Gaps 4;
 QY 523 GATCTCCAAGGAAGAAATTTGGGTGGTGTAGATTTTGTCAATAGTAGGAGTTATTCATAC 582
 DB 1 GATCTGAGGGCAAGGTCTATAGGCTGTACGACGGCGTCAACGGCAGGTGCGACCCCTAC 60
 QY 583 GATGACATGGATGGAACATCATGTAGCTTCAATAGCAGCTGTACTTGGAGCAGCAGT 542
 DB 61 GATGACGAGGACAGGAAACCCAGCTTCCGGGTATCTGTTCCGGAACCGCCGACCG 117
 QY 643 AATGGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGCGGGAATTAAGGTTCTAGGT 702
 DB 118 AATCCCAAGTACATAGGCGTCCGCGCGGAGCTGCTCGGCTCAAGGTTCCTCGGT 177
 QY 703 GCGGATGTTCTGGAGCANTATCTACTATAATTAAGGAGTTGAGTGGCCCTTGATTAAC 762
 DB 178 GCGGAGGTTCCGGAGCGTCTCCACCATCATCGCGGGTGTCTACTGGTCTCCAGAAC 237
 QY 763 AAGATAAGTACGAATTAAGGTCAATTAATCTTCTTCTTCTTCAAGCTCAGAGTCCGAC 822
 DB 238 AAGGACAAGTACGGGATMGGGTCAATCAACCTCTCCCTCGGCTCTCCAGAGTCCGAC 237
 QY 823 GGAACCGATCTCCCTCAGTCAAGCGCTCAACAGCGCTGGAGCGCGGTATAGTATGCTGC 882
 DB 298 GGAACCGATCTCCCTCAGTCAAGCGCTCAACAGCGCTGGAGCGCGGTATAGTATGCTGC 357
 QY 883 GTCCGCGCGGCAACAGCGGGCGGAAACCTTACACCTCGGGTCAACCGCGCGCGGAGC 942
 DB 358 GTCCGCGCGGCAACAGCGGGCGGAAACCTTACACCTCGGGTCAACCGCGCGCGGAGC 417
 QY 943 AAGGTCAATACCGTGGTGGTGTGACAGTGTGACAGCAACATCGCCAGCTTCTCCAGAGG 1002
 DB 418 AAGGTCAATACCGTGGTGGTGTGACAGTGTGACAGCAACATCGCCAGCTTCTCCAGAGG 477
 QY 1003 GGACGACCGGGAGGAAAGGCTCAAGCGGAGTGTGCGCGCGCGGTGACATCATATA 1062
 DB 478 GGACGACCGGGAGGAAAGGCTCAAGCGGAGTGTGCGCGCGGTGACATCATATA 537
 QY 1063 GCGCGCGCGGCAACAGCGGATGGCAACCGGATGAACAGTACTTCAACAAAGGCG 1122
 DB 538 GCGCGCGCGGCAACAGCGGATGGCAACCGGATGAACAGTACTTCAACAAAGGCG 597
 QY 1123 TGTGACCAAGCATGGCCACCCCGGCTTTCGGGGTGGCGGCTCATCTCCAGGCG 1182
 DB 598 TGTGATCCAGCATGGACACCCCGGCTTTCGGGGTGGCGGCTCATCTCCAGGCG 657
 QY 1183 CACCGAGCTGGACCCCGGACAAAGTGAAGCAGCGCTCATCGAGAGCGCGGATAGTC 1242
 DB 658 CACCGAGCTGGACCCCGGACAAAGTGAAGCAGCGCTCATCGAGAGCGCGGATAGTC 716
 QY 1243 GCGCCCAAGGATAGCGGACATCGCTACGGTGGGGTAGGGTGAACCTCTACAAGGCG 1302
 DB 717 GCGCCCAAGGATAGCGGACATCGCTACGGTGGGGTAGGGTGAACCTCTACAAGGCG 776
 QY 1303 ATCAAGTACGAGCTACGCGCAAGCTCACCTTCAAGGCTCGGCGCGGAGGAGG 1362
 DB 777 ATCAAGTACGAGCTACGCGCAAGCTCACCTTCAAGGCTCGGCGCGGAGGAGG 836

Thu Nov 7 10:12:16 2002

QY 1363 GCGACCCACACCTTCAGCTCAGCGGCGACCTTCGTGACCGCCACCCCTCTACTGGGAC 1422
DB 837 GGCA-CGACACCTTCGACGCGGGGCA-CTTGTGTGAAGCNCACCCCTCTNCTNGGAC 894
QY 1423 ACGG 1426
DB 895 ANGG 898
RESULT 10
ID AAV03761 standard; DNA; 507 BP.
AC AAV03761;
XX
DT 20-APR-1998 (first entry)
XX
DE Pyrococcus protease fragment coding sequence.
XX
KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX
OS Pyrococcus furiosus DSM-3638.
XX
FH Key Location/Qualifiers
FT CDS 2..505
FT /*tag= a
XX
XX WO9721823-A1.
XX
XX 19-JUN-1997.
XX
XX 07-NOV-1996; 95WO-JP03253.
XX
XX 12-DEC-1995; 95JP-0323285.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX Asada K, Kato I, Mitta M, Morishita M, Takakura H;
XX Tsunawasa S, Yamamoto K;
XX WPI: 1997-332794/30.
XX P-PSDB; AAW24127.
XX
XX Protease(s) and genes encoding them obtained from Thermococcus and
XX Pyrococcus strains - have extremely high thermal stability and are
XX useful industrially and as research reagents
XX
XX Disclosure; Page 118-119; 159pp; Japanese.
XX
XX This sequence represents a fragment of the coding sequence for the
XX protease from Pyrococcus furiosus DSM-3638 (see AAT85668 for full length
XX sequence). This sequence encodes a fragment of the protease of the
XX invention. The proteases of the invention have extremely high thermal
XX stability. The proteases can be used as research reagents, and
XX industrially in the food, drug and chemical industries.
XX
XX Sequence 507 BP; 152 A; 82 C; 136 G; 137 T; 0 other;
XX
XX Query Watch 25.6%; Score 507; DB 18; Length 507;
XX Best Local Similarity 100.0%; Pred. No. 3.1e-99;
XX Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 312 ACTAGTCATCTCAGGTTTAAACAGGGGTAAGCTTAAGCTTTTCAGGTGTTAGGTTTATCCA 371
DB 1 ACTAGTCATCTCAGGTTTAAACAGGGGTAAGCTTAAGCTTTTCAGGTGTTAGGTTTATCCA 60
QY 372 GGAAGACTACAAAGTTACAGTTTACAGCGAATTAGAGGACTGGATGCTGCAGCTCA 431
DB 61 GGAAGACTACAAAGTTACAGTTTACAGCGAATTAGAGGACTGGATGCTGCAGCTCA 120
QY 432 ASTTATGGCACTTACGTTTGGAACTTGGATATGATGTTCTGGAATCAATAGGAAT 491
DB 121 AGTTATGGCACTTACGTTTGGAACTTGGATATGATGTTCTGGAATCAATAGGAAT 180

QY 492 AATTGACACTGGAATTGACGCTTCTCATCCAGATCTCCAAGGAAAAGTAATTGGTGGGT 551
DB 191 AATTGACACTGGAATTGACGCTTCTCATCCAGATCTCCAAGGAAAAGTAATTGGTGGGT 240
QY 552 AGATTTTTCATGTTAGGAGTTATCCATACCATGACATGGACATCGAATCATGTAGC 611
DB 241 AGATTTTTCATGTTAGGAGTTATCCATACCATGACATGGACATCGAATCATGTAGC 300
QY 612 TTCAATAGCAGCTGTTACTGGAGCAGCAAGTAATGGCAAGTAATGGCAAGTAATGGCAAG 671
DB 301 TTCAATAGCAGCTGTTACTGGAGCAGCAAGTAATGGCAAGTAATGGCAAGTAATGGCAAG 360
QY 672 AGCTAAGCTGGCGGGAATTAAGGTTCTAGTGCCGATGGTTCTGGAAGCAATCTACTAT 731
DB 361 AGCTAAGCTGGCGGGAATTAAGGTTCTAGTGCCGATGGTTCTGGAAGCAATCTACTAT 420
QY 732 AATTAGGAGTTGAGTGGCGGTTGATACAAAGATAAGTACGGAATTAAGGTCATTAA 791
DB 421 AATTAGGAGTTGAGTGGCGGTTGATACAAAGATAAGTACGGAATTAAGGTCATTAA 480
QY 792 TCTTCT 818
DB 481 TCTTCT 507
RESULT 11
AAT08134
ID AAT08134 standard; DNA; 564 BP.
XX
XX AAT08134;
XX
XX 10-MAY-1996 (first entry)
XX
XX DNA sequence.
XX
XX Protease; hyperthermostable; thermostability; ss.
XX
XX Pyrococcus furiosus.
XX
XX WO9534645-A1.
XX
XX 21-DEC-1995.
XX
XX 05-JUN-1995; 95WO-JP01095.
XX
XX 26-JUL-1994; 94JP-0173912.
XX
XX 13-JUN-1994; 94JP-0130236.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX Asada K, Kato I, Mitta M, Morishita M, Tsunawasa S;
XX Yamamoto K;
XX WPI: 1996-049674/05.
XX
XX P-PSDB; AAR87011.
XX
XX Pyrococcus furiosus hyperthermostable protease gene - useful for
XX recombinant prodn. of hyperthermostable protease
XX
XX Example 4; Page 66-67; 85pp; Japanese.
XX
XX The invention relates to a gene (AAT08141) that codes for a
XX hyperthermostable protease (AAR87007) of Pyrococcus furiosus.
XX
XX 2 DNA sequences (AAT08133-34) are provided encoding peptides
XX (AAR87010-11).
XX
XX Sequence 564 BP; 121 A; 195 C; 161 G; 87 T; 0 other;
XX
XX Query Watch 21.6%; Score 427; DB 17; Length 564;
XX Best Local Similarity 86.1%; Pred. No. 4.5e-82;
XX Matches 485; Conservative 0; Mismatches 75; Indels 3; Gaps 1;

QY 595 CATGGAACTCATGACCTTCAATAGCAGCTGTGATGAGCAGCAAGTAATGGCAATAC 654
Db 4 CACGGAACCTCAGCTGGCGGAACAGTTCCGGAACAGCAGCG---TTAACTCCAGTAC 60
QY 655 AAGCGAATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGTCTAGGTGGCGATGTTCT 714
Db 61 ATAGCGGTGGCGCGCGGGAAGCTCGCTCGGTGTCAAGGTCTCGGTGGCGAGGTTCG 120
QY 715 GGAAGCATATCTACTATAATTAAGGAGCTGAGTGGCGGCTTGATTAACAAGATAAGTAC 774
Db 121 GGAAGCGCTCCACCATCATCGCGGTGTGACTGGGTCTGCGAAGCAAGATAAGTAC 180
QY 775 GGAATTAAAGCTCATTAATCTTCTGTGTTCAAGCCAGAGCTCGGAGCAACGACTCC 834
Db 181 GGGATAAGGCTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGAGGAGCGACTCC 240
QY 835 CTCAGTCAGCGCGCTCAACAACGCTGGAGCGCGGTATAGTCTGCTGCGCGCGCGC 894
Db 241 CTCAGTCAGCGCGCTCAACAACGCTGGAGCGCGGTATAGTCTGCTGCGCGCGCGC 300
QY 895 AACAGCGCGCGCTCAACAACGCTGGAGCGCGGTATAGTCTGCTGCGCGCGCGC 954
Db 301 AACAGCGCGCGCTCAACAACGCTGGAGCGCGGTATAGTCTGCTGCGCGCGCGC 360
QY 955 GTCGTCAGCTTCAGCAAGCAACATCGCCAGCTTCTCCAGCAGGAGCAACCGCGC 1014
Db 361 GTCGTCAGCTTCAGCAAGCAACATCGCCAGCTTCTCCAGCAGGAGCAACCGCGC 420
QY 1015 GACGGAAGGCTCAAGCGGAGTGTCTGCGCGCGCGGTGACATCATAGCGCGCGC 1074
Db 421 GACGGAAGGCTCAAGCGGAGTGTCTGCGCGCGCGGTGACATCATAGCGCGCGC 480
QY 1075 AGCGGAACCAAGCTGGGACCGCGGATTAAGCACTACTACCAAGCGCTCTGGAACCA 1134
Db 481 AGCGGAACCAAGCTGGGACCGCGGATTAAGCACTACTACCAAGCGCTCTGGAACCA 540
QY 1135 ATGCCACCGCGCGCTTTCGGG 1157
Db 541 ATGCCACCTCCCATGTTACCGG 563

RESULT 12

ID AAT85676 standard; DNA; 554 BP.
XX AAT85676;
AC AAT85676;
XX 20-APR-1998 (first entry)
XX Thermococcus protease fragment coding sequence.
XX Thermococcus protease fragment coding sequence.
XX Protease; research reagent; thermal stability; thermococcus celer; ss.
XX Thermococcus celer DSM-2476.
XX WO9721823-A1.
XX 19-JUN-1997.
XX 07-NOV-1996; 96WO-JP03253.
XX 12-DEC-1995; 95JP-032285.
XX (TAKI) TAKARA SHUZO CO LTD.
XX Asada K, Kato I, Mita M, Morishita M, Takakura H;
XX Tsunashima S, Yamamoto K;
XX WPI: 1997-332794/30.
XX P-PSDB: AAW24136.
XX Protease(s) and genes encoding them obtained from Thermococcus and
XX Pyrococcus strains - have extremely high thermal stability and are

PT useful industrially and as research reagents
XX Disclosure: Page 110-112; 159pp; Japanese.
XX This sequence represents a fragment of the coding sequence for the
CC protease from thermococcus celer DSM-2476 (see AAT85667 for full length
CC sequence). This sequence encodes a fragment of the protease of the
CC invention. The proteases of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries.

QY Sequence 564 BP; 121 A; 195 C; 161 G; 87 T; 0 other;

Query Match 21.6%; Score 427; DB 18; Length 564;
Best Local Similarity 86.1%; Pred. No. 4.5e-82;
Matches 485; Conservative 0; Mismatches 75; Indels 3; Gaps 1;

QY 595 CATGGAACTCATGACCTTCAATAGCAGCTGTGATGAGCAGCAAGTAATGGCAATAC 654
Db 4 CACGGAACCTCAGCTGGCGGAACAGTTCCGGAACAGCAGCG---TTAACTCCAGTAC 60
QY 655 AAGCGAATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGTCTAGGTGGCGATGTTCT 714
Db 61 ATAGCGGTGGCGCGCGGGAAGCTCGCTCGGTGTCAAGGTCTCGGTGGCGAGGTTCG 120
QY 715 GGAAGCATATCTACTATAATTAAGGAGCTGAGTGGCGGCTTGATTAACAAGATAAGTAC 774
Db 121 GGAAGCGCTCCACCATCATCGCGGTGTGACTGGGTCTGCGAAGCAAGATAAGTAC 180
QY 775 GGAATTAAAGCTCATTAATCTTCTGTGTTCAAGCCAGAGCTCGGAGCAACGACTCC 834
Db 181 GGGATAAGGCTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGAGGAGCGACTCC 240
QY 835 CTCAGTCAGCGCGCTCAACAACGCTGGAGCGCGGTATAGTCTGCTGCGCGCGCGC 894
Db 241 CTCAGTCAGCGCGCTCAACAACGCTGGAGCGCGGTATAGTCTGCTGCGCGCGCGC 300
QY 895 AACAGCGCGCGCTCAACAACGCTGGAGCGCGGTATAGTCTGCTGCGCGCGCGC 954
Db 301 AACAGCGCGCGCTCAACAACGCTGGAGCGCGGTATAGTCTGCTGCGCGCGCGC 360
QY 955 GTCGTCAGCTTCAGCAAGCAACATCGCCAGCTTCTCCAGCAGGAGCAACCGCGC 1014
Db 361 GTCGTCAGCTTCAGCAAGCAACATCGCCAGCTTCTCCAGCAGGAGCAACCGCGC 420
QY 1015 GACGGAAGGCTCAAGCGGAGTGTCTGCGCGCGCGGTGACATCATAGCGCGCGC 1074
Db 421 GACGGAAGGCTCAAGCGGAGTGTCTGCGCGCGCGGTGACATCATAGCGCGCGC 480
QY 1075 AGCGGAACCAAGCTGGGACCGCGGATTAAGCACTACTACCAAGCGCTCTGGAACCA 1134
Db 481 AGCGGAACCAAGCTGGGACCGCGGATTAAGCACTACTACCAAGCGCTCTGGAACCA 540
QY 1135 ATGCCACCGCGCGCTTTCGGG 1157
Db 541 ATGCCACCTCCCATGTTACCGG 563

RESULT 13

ID AAT85691 standard; DNA; 419 BP.
XX AAT85691;
AC AAT85691;
XX 20-APR-1998 (first entry)
XX Pyrococcus protease fragment coding sequence.
XX Pyrococcus protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX Pyrococcus furiosus DSM-3638.
XX Key Location/Qualifiers
FH

QY 572 GTTATCATACGATGACCATGGACATGGAACTCATGTAGCTTCAATAGCAGCTGGTACTG 631
 Db 1221 CGACCCCTTACCATGACGAGGACACGGAACCCAGCTTGCCTGATCGTTGGCGGAACG 1260
 QY 632 GACAGCAAGTAATGCGCAAGTACAAAGGAATGGCTCCAGSAGCTAAGCTGGCGGAATTA 691
 Db 1281 GGAGCG-...TTAACTCCCACTACATAGGCGTGGCGCCCGCGGCGAGCTCGTGGCGTCA 1337
 QY 692 AGGTTCTAGCTCCGATCGTCTGGAGGATATCTACTATATTAATTAAGGAGTGTAGTGG 751
 Db 1338 ASGTTCTCGTTCGACGCTGGGAGGCTCTCCACCATCATGCGGGGTGTGACTGGA 1397
 QY 752 CCCTTATCAACAAAGTAAGTACGGAATTAAGGTCATTAATCTTCTTCTGTTCAAGCC 811
 Db 1398 ACCTCAGAACTAGCAAGTACGGATAGGGTCAACCTCTCCCTCGGCTCCTGCC 1457
 QY 812 AGAGTC 818
 Db 1458 AGAGTC 1464

RESULT 15

ABL54900
 ID ABL54900 standard; DNA; 2121 BP.
 AC ABL54900;
 DT 31-MAY-2002 (first entry)
 XX T. yonsei subtilisin-like serine protease coding sequence.
 DE Subtilisin-like serine protease; ss.
 KW Thermoaerobacter yonsei.
 XX
 OS Thermoaerobacter yonsei.
 FH Key Location/Qualifiers
 FT CDS 142..1779
 FT /*tag= a
 FT /product= "subtilisin-like serine protease"
 XX
 PN KR2000072141-A.
 XX
 XX 05-DEC-2000.
 XX
 XX 04-AUG-2000; 2000KR-0045411.
 XX
 XX 04-AUG-2000; 2000KR-0045411.
 XX
 XX (KIMY/) KIM Y S.
 XX
 XX Chang HJ, Kim DH, Byun YR, Kim YS;
 XX WPI; 2001-298092/31.
 XX P-PSDB; ABB09483.
 XX
 XX New DNA sequence of thermophilic protein decomposition enzyme and
 XX protein derived therefrom -
 XX
 XX Claim 1; Page 6; 15pp; Korean.
 XX
 CC This sequence represents the DNA encoding the Thermoaerobacter
 CC yonsei subtilisin-like serine protease of the invention.
 XX
 SQ Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 20 other;

Query Match 9.98; Score 195; DB 23; Length 2121;
 Best Local Similarity 61.5%; Pred. No. 3.4e-32;
 Matches 372; Conservative 0; Mismatches 215; Indels 18; Gaps 3;
 QY 478 ATCAATAGGAATTAATGACACTGCAATGAGCTTCTCATCCAGATCTCCAA---CGA 534
 Db 490 ATAACAATAGCAATATTGACACAGGTATAGACGGGAATACGTTGACCTCTCAGGTGGA 549

QY 535 AAAGTAATTGGTGGGTAGATTTGTCAATGGTAGGAGCTTATCCATAGATGACCATGGA 594
 Db 550 AAAATAATAGGATGGAAGAACTTTATCAACAACAAANATACACCATAGCAGCAATGGC 609
 QY 595 CATGGAACCTCATGTAGCTTCAATAGCAGCTGGTACTGGACGAGCAAGTAATGGCAAGTAC 654
 Db 610 CATGGAACCTCAGCTAGCAAGTATTGCTGCAAGGTACAGCTGCTG---GAAACAGTCTTTAC 666
 QY 655 AAGGGAATGCTCCAGGAGCTAAGCTGGCGGAATTAAGCTTCTAGGTGCGGATGGTTCT 714
 Db 667 AAGCGCTTCTCTCTGATGCTTTGTTGGTAAAGTAAAGTTTAGATGCAAAATGGAAGC 726
 QY 715 GGAAGCATATCTACTATATTAATTAAGGAGTTGAGTGGCGGCTTGATTAACAAGATAAGTAC 774
 Db 727 GGCAGCATGAGCACTGTAACTGCAGGAATTAAGTGGGCTGTTCAAAATAAAGATGTATAC 786
 QY 775 GGAATTAAGGTCAATTAATCTTCTTCTTGGTTCAAGCCAGAGCTCCGACGGAACCGACTCC 834
 Db 787 GGAATCAAGTTATAAATTTAAGCTCGGCACCTTCTACAAGTTCTGATGGAAGTACTCT 846
 QY 835 CTCAGTCAGCGCTCAACACGCTCGGAGCGCGGTATATAGTCTCGTGGCGCGCGC 894
 Db 847 ACCTATTAGCAGTGAATAGGACAGTAGATAGCGGTATTGTAGTGTAGCAGCAGCA 906
 QY 895 AACAGCGGCGGACACACCTACACCGTGGCTCACCCGCGCGGAGCAGAGGTCTATAACC 954
 Db 907 AACTCTGGCCCTGCAAAATACACCATAGGCTCCCTGGTGTGCGGAAAGCCATACA 966
 QY 955 GTC-----GGTGCAGTTGACAGCAACAGCAACATCGCAGCTCTCCAGCAGG 1002
 Db 967 GTCGAGCAATGGCAGATCTAGGTGAACTTGGCTTTAACCTTGCAGCTTTTCCAGCGC 1026
 QY 1003 GGACCCAGCGCGGAGGAGCTCAAGCGGAGTGTGTCGCCCCCGCGCTTGACATCAT 1062
 Db 1027 GGTCTACTCTGACGGGAAGTAATAAACCTGACATTGCGCCCCCAGGATATAATAACT 1086
 QY 1063 GCCCC 1087
 Db 1087 GCCGC 1091

Search completed: November 2, 2002, 08:28:26
 Job Line : 294 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 2, 2002, 10:33:33 ; Search time 51.5 seconds
(without alignments)
12789.264 Million cell updates/sec

Title: US-09-841-553-6

Perfect score: 1977

Sequence: 1 ATGAAGGGCTGAAGCTCT.....AGCGCGCTCTACTACGGG 1977

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 165577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Databases :

- 1: /cgn2_5/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_5/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_5/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_5/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_5/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_5/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_5/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 8: /cgn2_5/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 9: /cgn2_5/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_5/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_5/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_5/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_5/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_5/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1360.4	69.1	1962	12	US-10-090-624-15
2	1360.8	68.8	1977	12	US-10-090-624-11
3	837	42.3	1236	12	US-10-090-624-11
4	104.6	5.3	1635	10	US-09-884-761-20241
5	104.6	5.3	1973	10	US-09-864-761-3471
6	97.8	4.9	1075	10	US-09-864-761-19241
7	97.8	4.9	1403	10	US-09-864-761-2513
8	97.6	4.9	2712	10	US-09-748-033-4
9	84.8	4.3	390	10	US-09-790-399-7
10	83.4	4.2	1185	10	US-09-887-576-784
11	83.2	4.2	15720	10	US-09-922-217-1058
12	83.2	4.2	1329	10	US-09-833-263-1058
13	82.6	4.1	1329	10	US-09-974-300-1334
14	80.4	4.1	863	10	US-09-974-300-5249
15	79.2	4.0	1107	10	US-09-748-033-5
16	71.8	3.6	867	10	US-09-216-393-340
17	71.8	3.6	867	10	US-09-216-393-342
18	71.8	3.6	1294	10	US-09-748-033-2
19	71.8	3.6	1397	10	US-09-216-393-343

C	20	71.8	3.6	1397	10	US-09-216-393-345	Sequence 345, Appl
	21	70	3.5	905	10	US-09-748-033-5	Sequence 5, Appl
	22	69.6	3.5	2352	10	US-09-476-242-26	Sequence 26, Appl
	23	69.2	3.5	1128	9	US-09-934-060A-27	Sequence 27, Appl
	24	69.2	3.5	1769	9	US-09-934-060A-12	Sequence 12, Appl
	25	69	3.5	2517	10	US-09-476-242-16	Sequence 16, Appl
	26	69	3.5	2523	10	US-09-476-242-15	Sequence 15, Appl
	27	68.6	3.5	1518	9	US-09-934-060A-23	Sequence 23, Appl
	28	68.6	3.5	1518	9	US-09-934-060A-29	Sequence 29, Appl
	29	68.6	3.5	1668	9	US-09-934-060A-5	Sequence 5, Appl
	30	68.6	3.5	2159	9	US-09-934-060A-1	Sequence 1, Appl
	31	68.6	3.5	2159	9	US-09-934-060A-3	Sequence 3, Appl
	32	67.6	3.4	2358	10	US-09-476-242-25	Sequence 25, Appl
	33	67.6	3.4	2517	10	US-09-476-242-17	Sequence 17, Appl
	34	67.6	3.4	2529	10	US-09-476-242-14	Sequence 14, Appl
	35	67.6	3.4	2359	10	US-09-476-242-13	Sequence 13, Appl
	36	67.6	3.4	2541	10	US-09-476-242-9	Sequence 9, Appl
	37	67.6	3.4	2541	10	US-09-476-242-10	Sequence 10, Appl
	38	67.6	3.4	2541	10	US-09-476-242-11	Sequence 11, Appl
	39	67.6	3.4	2541	10	US-09-476-242-12	Sequence 12, Appl
	40	66.6	3.4	1371	10	US-09-350-756-3	Sequence 3, Appl
	41	66.4	3.4	4689	10	US-09-861-288-34	Sequence 34, Appl
	42	66.4	3.4	36778	10	US-09-861-288-5	Sequence 5, Appl
C	43	66	3.3	446	10	US-09-861-288-5	Sequence 20699, A
	44	65.8	3.3	1191	10	US-09-815-242-7694	Sequence 7694, Ap
	45	64.2	3.2	2298	10	US-09-476-242-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-10-090-624-15
; Sequence 15, Application US/10090624
; Patent No. US2002013233A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Inuoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-05
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-15

Query Match 69.1%; Score 1366.4; DB 12: Length 1962;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 1605; Conservative 0; Mismatches 356; Indels 15; Gaps 1;

Oy 1 ATCAAGGGCTGAAGCTCTCATATTAGTATTGTTTCTAGTTAGTAGGAGC 60
Db 1 ATGAAGGGCTGAAGCTCTCATATTAGTATTGTTTCTAGTTAGTAGGAGC 60
Oy 61 GTAGCGGAGCTCCAGAGAGAACTTCAACAAGTAAGAAATGTTGACAGAACTATGCT 120
Db 61 GTAGCGGAGCTCCAGAGAGAACTTGAACAAGTANGAATGTTGAGAGAACTATGCT 120
Oy 121 CTGCTAAGCGGAGCTGTTGAGAAATTCAGAAATTCAGAAATTCAGAAATTCAGC 180
Db 121 CTGCTAAGCGGAGCTGTTGAGAAATTCAGAAATTCAGAAATTCAGAAATTCAGC 180

Thu Nov 7 10:12:18 2002

121 CTGCTAACGCCAGGAGCTGTTCAGAAAAATTCAAAATTCGATCTCTACGAGGAATCAGC 180
Db
181 ACAGTAATTTGTTTGAAGACATAGGGAAGAAATTCAGTAGAGTTCTTCAGTTA 240
Qy
181 ACAGTAATTTGTTTGAAGACATAGGGAAGAAATTCAGTAGAGTTCTTCAGTTA 240
Db
241 ATGGCTGCAAAAGTATAGTATGTACCATATTATACCGGCAATAGTGGCCATCTTAAG 300
Qy
241 ATGGCTGCAAAAGTATAGTATGTACCATATTATACCGGCAATAGTGGCCATCTTAAG 300
Db
301 GTTAGAGACTTACTAGTATCTCAGGTTTAAACAGGGGTAAAGTAAAGCTTTTCAGGTGT 360
Qy
301 GTTAGAGACTTACTAGTATCTCAGGTTTAAACAGGGGTAAAGTAAAGCTTTTCAGGTGT 360
Db
361 AGGTTTATCCAGGAAGACTACAAAGTATACAGTTTCAAGAGTATAGAGACTTGGATGAG 420
Qy
361 AGGTTTATCCAGGAAGACTACAAAGTATACAGTTTCAAGAGTATAGAGACTTGGATGAG 420
Db
421 TCTGCAGCTCAAGTTATGGCACTTACGTTTGGAACTTGGGATATGATGTTCTGGAATC 480
Qy
421 TCTGCAGCTCAAGTTATGGCACTTACGTTTGGAACTTGGGATATGATGTTCTGGAATC 480
Db
481 ACATAGGAAATTTGACACTTGAATTTGACGTTTCTCATCCAGATCTCAAGGAATCTA 540
Qy
481 ACATAGGAAATTTGACACTTGAATTTGACGTTTCTCATCCAGATCTCAAGGAATCTA 540
Db
541 ATTGGGTGGGTAGATTTTGTCAATGTAGGAGTTATCATCATGATGACCATGGATGGA 600
Qy
541 ATTGGGTGGGTAGATTTTGTCAATGTAGGAGTTATCATCATGATGACCATGGATGGA 600
Db
601 ACTCATCTAGCTTCAATAGCAGTGTGTTGAGCAGCAAGTATGGCAAGTACAAAGGGA 660
Qy
601 ACTCATCTAGCTTCAATAGCAGTGTGTTGAGCAGCAAGTATGGCAAGTACAAAGGGA 660
Db
661 ATGGCTCCAGAGCTAAGCTGGCGGGAATTAAGGTTCTAGTGCCGATGTTCTGGGAGC 720
Qy
661 ATGGCTCCAGAGCTAAGCTGGCGGGAATTAAGGTTCTAGTGCCGATGTTCTGGGAGC 720
Db
721 ATATCTACTATTAATTAAGGAGTTCAGTGGCGGTTGATTAACAAGATTAAGTACGAAT 780
Qy
721 ATATCTACTATTAATTAAGGAGTTCAGTGGCGGTTGATTAACAAGATTAAGTACGAAT 780
Db
781 AAGGTCATTAACTTTCTTCTGTTCAAGCCAGCTCCGAGCGGAGCTCCCTCAST 840
Qy
781 AAGGTCATTAACTTTCTTCTGTTCAAGCCAGCTCCGAGCGGAGCTCCCTCAST 840
Db
841 CAGGCGTCAACAACGCTGGGACCGGATATAGTCTGGTCCGCGCGGCGGACAGC 900
Qy
841 CAGGCGTCAACAACGCTGGGACCGGATATAGTCTGGTCCGCGCGGCGGACAGC 900
Db
901 GGCGGACACCTACACCTGCTCCGCTCACCGCCCGCGGAGCGTCTATAACCTCGGT 960
Qy
901 GGCGGACACCTACACCTGCTCCGCTCACCGCCCGCGGAGCGTCTATAACCTCGGT 960
Db
961 GCAGTTGACAGCAACGACAACTTCGCGAGCTTCTCAGAGGGGACCCAGCGGAGCGA 1020
Qy
961 GCAGTTGACAGCAACGACAACTTCGCGAGCTTCTCAGAGGGGACCCAGCGGAGCGG 1020
Db
1021 AGGCTACGCGGAGTGTGTCGCGCGCGGCTTGACATCATAGCCCGCGCGGCGGACGGA 1080
Qy
1021 AGGCTACGCGGAGTGTGTCGCGCGCGGCTTGACATCATAGCCCGCGCGGCGGACGGA 1080
Db
1081 ACCAGCATGGGACCCCATTAACGACTACTACACGAGGCTCTGGAAACCCAGCATGGCC 1140
Qy
1081 ACTAGCATGGGTCACCCATTAATGACTATTACAGCAGCTCTCGGACATCAATGCCA 1140
Db
1141 ACCCGGACGCTTTTCGGGCGTTGGCGGCTCTATCCAGCCAGCCAGGAGGACCCG 1200
Qy
1141 ACTCTCAGCTAGCTGGTATTTGACAGCCCTCTTGTCTCCAGCAGCAGCCAGCTGCA 1200
Db
1201 GACAGGTGAAGACCCCTCTCATCGAGACCGCGGACATAGTCGCCCCCAAGGATGCG 1260
Qy
1201 GACAGGTGAAGACCCCTCTCATAGAACTGCTGATATCGTAAAGCCAGCATGAATGCC 1260
Db

1261 GACATCGCTACGCTAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAGGACTAC 1320
Db
1261 GATATAGCTACGCTAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAGGACTAC 1320
Qy
1321 GCCAAGCTCAGCTTACGCTAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAGG 1380
Qy
1321 GCCAAGCTCAGCTTACGCTAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAGG 1380
Db
1381 GTACAGGCGGCACCTCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGC 1440
Qy
1381 GTACAGGCGGCACCTCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGC 1440
Db
1441 GATCTTACCTCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAGGCTTAC 1500
Qy
1441 GATCTTACCTCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAGGCTTAC 1500
Db
1501 GGCTTCAGAGGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1560
Qy
1501 GGCTTCAGAGGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1560
Db
1561 AGCTACAGGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1620
Qy
1561 AGCTACAGGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1620
Db
1621 TCCGCGGCGGCAACCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1680
Qy
1621 TCCGCGGCGGCAACCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1680
Db
1681 CAGACTTACCGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1740
Qy
1681 CAGACTTACCGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1740
Db
1741 GTCACAGGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1800
Qy
1741 GTCACAGGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1800
Db
1801 CTGACGCTTACCTCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1860
Qy
1801 CTGACGCTTACCTCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1860
Db
1861 AGCTACAGGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1920
Qy
1861 AGCTACAGGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1920
Db
1921 GCTTACAGGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1976
Qy
1921 GCTTACAGGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1976
Db
1976 GCTTACAGGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1976
Qy
1976 GCTTACAGGCTGCTGCTAGCAGGCTAGGCTGAACGCTCTACAGGCCATCACTAGCAG 1976
Db

RESULT 2
US-10-090-624-11
; Sequence 11, Application us/10090624
; Patent No. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyoko
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-5
; CURRENT APPLICATION NUMBER: US/10/090, 624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-10-090-624-11

Query Match 68.8%; Score 1360.8; DB 12; Length 1977;
Best Local Similarity 83.9%; Pred. No. 0;
Matches 1573; Conservative 0; Mismatches 292; Indels 9; Gaps 3;

OY 107 AGAAGAACTATGCTCTGCTTAACCCGAGGACTGTTTCAGAAAATTTCAAAATTTGAATCCTA 166
DB 110 AGAAGAACTATGCTCTGCTTAACCCGAGGACTGTTTCAGAAAATTTCAAAATTTGAATCCTA 166
OY 167 ACAGAGAAATACGACAGTAAATTTGTAATTTGAAACCATACGGAAGAAATTTCAAGTAA 226
DB 170 ACCGAGAAATACGACAGTAAATTTGTAATTTGAAACCATACGGAAGAAATTTCAAGTAA 226
OY 227 GAGTTCTGAGTTAATGGGTGCAAAAGTTAGGTATGTTAGTATGTTACCATATTTATACCGCAATAG 286
DB 230 AGGTACTGAGGCTCATGGGCGCCAGGTCAGTAACTCTCAAGATATCTCTCAAGATATCTCTGCTGCTG 289
OY 287 CTGCGGATCTTAAAGTTAGAGACTTACTAGTCACTCTCAGGTTTACAGGGGGTAAAGCTA 346
DB 290 CGGTTAAATAAAGGCGGACCTCTGCTGATCGGGGCGATGATA---GACACGGGTT 346
OY 347 AGCTTTACAGTGTAGTTTATCCAGGAAGACTACAAAGTTTACAGTTTACAGCAAGATTAG 406
DB 347 ACTTCGGTACACAAAGGGTCTCGGCATAAAGTTTACAGGAAGATTACAAAGTTTCAGG 405
OY 407 AAGGACTGGATGAGTCTGCAAGTCAAGTTATGGCAACTTACGTTTGGAACT---TGGGAT 463
DB 407 TTGACGAGCGCACTTCGCTCTCCAGATAGGGCGGATACGCTCTGGAATCTCCCTCGCT 466
OY 464 ATGATGTTCTGGAATACAAATAGGAATTAATTCAGACTGGAATTCAGCGCTTCTCATCCAG 523
DB 467 ACAGCGAAGCGGTGGTGGTTCGCTATCGTCTGATAGCGGTTATAGCGGGAACCAACCCG 526
OY 524 ATCTCCAGGAAGCAATTAATGGTGGTGTAGTTTGTCAATGGTATAGAGTTATCCATACG 583
DB 527 ATCTGAGGCAAGTCTATAGGCTGTTGACGAGCGGCTCAAGCGAGGCTGACCGCCCTACG 586
OY 584 ATGACATGACATGGAATCATGTAGCTTCAATAGCAGCTGCTGCTGCGGAACCGGACG---TTA 643
DB 587 ATGACATGACATGGAATCATGTAGCTTCAATAGCAGCTGCTGCTGCGGAACCGGACG---TTA 643
OY 644 ATGCAAGTACAGGGAATGGCTCCAGGAGCTAAGCTGCGGCGGCAATTAAGTTCTAGGTG 703
DB 644 ACTCCAGTACATAGGCTGCGGCGGCAATTAAGTTCTAGGTG 703
OY 704 CCGATGTTCTGGAAGCATATCTATTAATTAAGGAGTGTAGTGGGCGGCTTGTATACCA 763
DB 704 CCGAGGTTGCGGAAGGCTTCCACCATCATCGCGGTTGTAGTGGGCTGCTGCGGCAACA 763
OY 764 AAGATAGTACGGAATTAAGTGTCAATTAATTAAGGAGTGTAGTGGGCGGCTTGTATACCA 823
DB 764 AGGACAGTACGGAATTAAGTGTCAATTAATTAAGGAGTGTAGTGGGCGGCTTGTATACCA 823
OY 824 GAACGACTCTCTAGTACGCGCTCAACAACGCTTCCAGCGCGGCTTGTATAGTGTGCG 883
DB 824 GAACGACTCTCTAGTACGCGCTCAACAACGCTTCCAGCGCGGCTTGTATAGTGTGCG 883
OY 884 TCAGCGCGGCAACAGCGGCGGCAACACTACACCTGCGGCTTCCAGCGCGGCTTGTATAGTGTGCG 943
DB 884 TCAGCGCGGCAACAGCGGCGGCAACACTACACCTGCGGCTTCCAGCGCGGCTTGTATAGTGTGCG 943
OY 944 AGGTATACGCTCGGTGCAAGTGTGACAGCAACGACATATCCAGCTTCTCCAGCAGG 1003
DB 944 AGGTATACGCTCGGTGCAAGTGTGACAGCAACGACATATCCAGCTTCTCCAGCAGG 1003
OY 1004 GACCAACCGGAGCGGAGGCTCAAGCGGAGTCTGTCGCGCGCGGCTTGTATAGTGTGCG 1063
DB 1004 GACCAACCGGAGCGGAGGCTCAAGCGGAGTCTGTCGCGCGCGGCTTGTATAGTGTGCG 1063

OY 1064 CCCCGCGGCCAGCGGAACCCAGCATGGGACCCCGATTAAGCACTACTACACCAAGCCCT 1123
DB 1064 CCCCGCGGCCAGCGGAACCCAGCATGGGACCCCGATTAAGCACTACTACACCAAGCCCT 1123
OY 1124 CTGGAACCAAGCATCGCCACCCCGCAGCTTTCTGGGCGGTTGGCGGCTCATCTCCAGGCC 1183
DB 1124 CTGGAACCAAGCATCGCCACCCCGCAGCTTTCTGGGCGGTTGGCGGCTCATCTCCAGGCC 1183
OY 1184 ACCCGAGCTGACCGCCGCGAAGAGTGAAGACCCCGCTCATCGAGACCGCGGACATAGTCG 1243
DB 1184 ACCCGAGCTGACCGCCGCGAAGAGTGAAGACCCCGCTCATCGAGACCGCGGACATAGTCG 1243
OY 1244 CCCCAAGGATAGGCGACATCGCCTACGCTACGCTACGCTAGGCTAGGCTGACGCTCTCA 1303
DB 1244 CCCCAAGGATAGGCGACATCGCCTACGCTACGCTAGGCTAGGCTGACGCTCTCA 1303
OY 1304 TCACTACGACGACTACGCGCAAGCTCACTTACGCGGCTTCGCGGCAAGGAGCG 1363
DB 1304 TCACTACGACGACTACGCGCAAGCTCACTTACGCGGCTTCGCGGCAAGGAGCG 1363
OY 1364 CCACCCACGCTTCGAGCTACGCGCGCCACCTTGGTGAAGCCGCCACCTCTACTGGGACA 1423
DB 1364 CCACCCACGCTTCGAGCTACGCGCGCCACCTTGGTGAAGCCGCCACCTCTACTGGGACA 1423
OY 1424 CGGCTCGAGCGACATCGACCTCTACCTCTAGGACCCCAAGCGGAGGAGTTGACTACT 1483
DB 1424 CGGCTCGAGCGACATCGACCTCTACCTCTAGGACCCCAAGCGGAGGAGTTGACTACT 1483
OY 1484 CCTACACCGCTACTACGCTTCGAGAAAGTGGCTACTACACCGGAGGAGTTGACTACT 1543
DB 1484 CCTACACCGCTACTACGCTTCGAGAAAGTGGCTACTACACCGGAGGAGTTGACTACT 1543
OY 1544 GGAGGCTCAGCTTCGCTAGCTACAGGGCGCGGCAAGCTACCAAGTCTGAGCTGCTCAGCG 1603
DB 1544 GGAGGCTCAGCTTCGCTAGCTACAGGGCGCGGCAAGCTACCAAGTCTGAGCTGCTCAGCG 1603
OY 1604 ACAGGAGCTCAGCTTCGCGGCGGCAAGCTACCAAGTCTGAGCTGCTCAGCG 1663
DB 1604 ACAGGAGCTCAGCTTCGCGGCGGCAAGCTACCAAGTCTGAGCTGCTCAGCG 1663
OY 1664 CCGGACACCGGACACACCTTCCAGCGGCTTCCAGCGGTTCCGTTAAGGACTTGGGACCAAGCG 1723
DB 1664 CCGGACACCGGACACACCTTCCAGCGGCTTCCAGCGGTTCCGTTAAGGACTTGGGACCAAGCG 1723
OY 1724 ACACCTTCACCATGAGCTCAGCTACAGCGGTTGCCACCAAGATTAAGCTGACCTTTCG 1783
DB 1724 ACACCTTCACCATGAGCTCAGCTACAGCGGTTGCCACCAAGATTAAGCTGACCTTTCG 1783
OY 1784 ATACTTCTTACAGCTCAGCTTCTACCTCTACGACCCCAAGCAACCTCGTTGACA 1843
DB 1784 ATACTTCTTACAGCTCAGCTTCTACCTCTACGACCCCAAGCAACCTCGTTGACA 1843
OY 1844 GGTCCACGTCGCGGCAAGCTACGAGCAGCTGAGTACCGCAACCCCGCGGAGCT 1903
DB 1844 GGTCCACGTCGCGGCAAGCTACGAGCAGCTGAGTACCGCAACCCCGCGGAGCT 1903
OY 1904 GAGCTTCTCGCTGAGCTACAGCTTACGCTTACGCTGCGGCTGAGCTCAAGCGCG 1963
DB 1904 GAGCTTCTCGCTGAGCTACAGCTTACGCTTACGCTGCGGCTGAGCTCAAGCGCG 1963
OY 1964 TCGTCTACTAGCGG 1977
DB 1964 TCGTCTACTAGCGG 1977

RESULT 3
US-10-090-624-2
; Sequence 2, Application US/10090624
; Patent No. US2002013335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, HIKARU
; APPLICANT: MORISHIITA, MIKIO
; APPLICANT: SHIMOJO, TOMOKO

Thu Nov 7 10:12:18 2002

APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,524
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 1236
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-10-090-524-2

Query Match 42.1%; Score 837; DB 12; Length 1236;
Best Local Similarity 80.1%; Pred. NO. 1.3e-211;
Matches 984; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

QY 397 GCAGAAATAGAGGACTGGATGCTGCAGCTCAAGTTATGGAACCTTACGCTTTGGAAC 456
DB 1 GCAGAAATAGAGGACTGGATGCTGCAGCTCAAGTTATGGAACCTTACGCTTTGGAAC 60
QY 457 TTGGGATATGATGGTTCTGGAATCACAATAGGAATTAATGACACTGGAATTCAGCTTCT 516
DB 61 TTGGGATATGATGGTTCTGGAATCACAATAGGAATTAATGACACTGGAATTCAGCTTCT 120
QY 517 CATCCAGATCTCCAGGAAAGTAATTTGGGTGGCTAGATTTTGTCAATGCTAGGAGTTAT 576
DB 121 CATCCAGATCTCCAGGAAAGTAATTTGGGTGGCTAGATTTTGTCAATGCTAGGAGTTAT 180
QY 577 CCATACGATGACATGGATGGAATCTAGCTTCAATAGCAGCTGGTACTGAGCA 636
DB 181 CCATACGATGACATGGATGGAATCTAGCTTCAATAGCAGCTGGTACTGAGCA 240
QY 637 GCAAGTAATGCAAGTACAAGGAATGGCTCCAGGAGCTTAGCTGGCGGAATTAAGTT 696
DB 241 GCAAGTAATGCAAGTACAAGGAATGGCTCCAGGAGCTTAGCTGGCGGAATTAAGTT 300
QY 697 CTAGTGGCGATGTTCTGGAAGCATATCTACTATAATTAAGGAGTTAGTGGCGGTT 756
DB 301 CTAGTGGCGATGTTCTGGAAGCATATCTACTATAATTAAGGAGTTAGTGGCGGTT 360
QY 757 GATACAAAGTAAAGTACGGAATTAAGTCTATTATCTTCTTCTTCTTCAAGCCAGAGC 816
DB 361 GATACAAAGTAAAGTACGGAATTAAGTCTATTATCTTCTTCTTCTTCAAGCCAGAGC 420
QY 817 TCCAGCGGAACGACTCCCTCAGTACGCGCTCAACAGCGCTGGGAGCGGCTATAGTA 876
DB 421 TCAGATGGTACTAGCTCTAAGTCAAGGCTTAAAGTCTATTATCTTCTTCTTCAAGCCAGAGC 480
QY 877 GTCTGGTCCCGCGGCAACAGCGGCGGCAACAGCTTACAGCTGGGCTCCAGCGCGCC 936
DB 481 GTCTGGTCCCGCGGCAACAGCGGCGGCAACAGCTTACAGCTGGGCTCCAGCGAGCT 540
QY 937 GCGAGAGAGGTATACCGCTGGTGGAGTTGACAGCAACAGCAACATCCAGCTTCTCC 996
DB 541 GCAAGCAAGTATTAACAGTTGGAGCGGTTGACAGTATGATTTATAACAGCTTCTCA 600
QY 997 AGCAGGGAGCCGACCGGAGGAGGCTCAAGCGGAGTGTCTGCGCCCGCGGCTTGAC 1056
DB 601 AGCAGGGAGCCGACCGGAGGAGGCTCAAGCGGAGTGTCTGCGCCCGCGGCTTGAC 660
QY 1057 ATCATAGCCCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1116
DB 661 ATCATAGCCCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 720
QY 1117 AAGGCGCTTGGAAACCAAGCAATGCCACCCGAGTTTCGGGGGCTTGGCGGCTCATCTC 1175

DB 721 GCAGCTCTGGGACATCAATGGCAACTCTCTCAGTAGTGGTATTGACGCCCTCTTGCTC 780
QY 1177 CAGGCCCGCCGAGCTGCAGCCGGCAAGGTGAAGACCGCCCTCATCGAGACCGCCGAC 1236
DB 781 CAAGCACACCCGAGCTGGCTCCAGCAAGTAAACAGCCCTCATAGAACTGCTGAT 840
QY 1237 ATAGTCCGCCCCCAAGGAGATAGCGGACATCGCTACGGTGGGGTAGGTGAACGCTTAC 1296
DB 841 ATCGTAAAGCCAGATGAATAGCCGATATAGCTTACGGTGGAGTAGGTAAATGATAC 900
QY 1297 AAGGCAATCAAGTACGAGACTAGCCCAAGCTCACCTTACCGGCTCCGTCCCGGACAG 1356
DB 901 AAGGCTATAACTAGCAATAACTATGCAAGCTAGTGTCTACTGGATATGTGGCAACAA 960
QY 1357 GGAAGCGCCACCCACACTTTCAGCTCAGCGGCGCCACCTTCGTGACCGCCCTCTTAC 1416
DB 961 GCGACCAAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTCTTAAGTGGCAATGAT 1020
QY 1417 TGGGACCGGCTCGAGGACATCGACCTTACCTTACGACCCCAAGCGGAGAGGTT 1476
DB 1021 TGGGACAAATGCAATAGCGACCTTATCTTACCTTACGATCCCAATGGAACAGGTT 1080
QY 1477 GACTACTCTACACCGCTACTACGGCTTCGAGAGGCTTCGAGAGCTTACAGCGGAGCT 1536
DB 1081 GACTACTCTACACCGCTACTATGATGATGCAAGGTTGGTTATTATACCCCACTGAT 1140
QY 1537 GGAAGTGGAGGTCAGGTCGCTCAGCTACAGGCGCGGCGGCAACTTACAGCTCGAGCT 1596
DB 1141 GGAAGTGGAGGTCAGGTCGCTCAGCTACAGGCGCGGCGGCAACTTACAGCTCGAGCT 1200
QY 1597 GTCAGGCGGAGGCTCAGCCAGCTCCG 1625
DB 1201 GTAAGTATGCTTCCCTTTCACAGCGCTGG 1229

RESULT 4
US-09-854-761-20241
Sequence 20241, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aemica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663


```

1 PRIOR APPLICATION NUMBER: PCT/US01/006565
2 PRIOR FILING DATE: 2001-01-30
3 PRIOR APPLICATION NUMBER: PCT/US01/006568
4 PRIOR FILING DATE: 2001-01-30
5 PRIOR APPLICATION NUMBER: PCT/US01/006653
6 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/006662
8 PRIOR FILING DATE: 2001-01-30
9 PRIOR APPLICATION NUMBER: PCT/US01/006661
10 PRIOR FILING DATE: 2001-01-30
11 PRIOR APPLICATION NUMBER: PCT/US01/006670
12 PRIOR FILING DATE: 2001-01-30
13 PRIOR APPLICATION NUMBER: US 60/234,687
14 PRIOR FILING DATE: 2000-09-21
15 PRIOR APPLICATION NUMBER: US 09/608,408
16 PRIOR FILING DATE: 2000-06-30
17 PRIOR APPLICATION NUMBER: US 09/774,203
18 PRIOR FILING DATE: 2001-01-29
19 NUMBER OF SEQ ID NOS: 49117
20 SOFTWARE: Annonxam Sequence Listing Engineering
21 SEQ ID NO: 3471
22 LENGTH: 1973
23 TYPE: DNA
24 ORGANISM: Homo sapiens
25 FEATURE: MAP TO ACC06547.9
26 OTHER INFORMATION: EXPRESSED IN LUNG; S
27 OTHER INFORMATION: EXPRESSED IN PLACENT
28 OTHER INFORMATION: EXPRESSED IN HEA
29 OTHER INFORMATION: EXPRESSED IN HEAT
30 OTHER INFORMATION: EXPRESSED IN HEART
31 OTHER INFORMATION: EXPRESSED IN HBL100
32 OTHER INFORMATION: EXPRESSED IN BT474
33 OTHER INFORMATION: EXPRESSED IN TETAL
34 OTHER INFORMATION: EXPRESSED IN ADULT
35 OTHER INFORMATION: EXPRESSED IN BONE M
36 OTHER INFORMATION: EXPRESSED IN BRAIN,
37 ITS-09-864-761-3471

```

Query Match	5.3%	Score 104.6	DB 10	Length 1373
Best Local Similarity	44.6%	Pred. No. 6.7e-10	Indels	Gaps
Matches 499	Conservative	0	Mismatches 614	
QY	814	AGCTCCGACGGAAACGACTCCCTCTAGTCAGCGCGTCAACAGCGCTGGACCGCCGGTATA	873	
DB	570	ATCACCATACACCAACCCATCATACCATCACTACCAACCAACCAACCACTCACTACC	629	
QY	874	GTATGTCGTGCGCGCGGGGACACCGGGCGGACACTACAGCTGGCTGGCTCAACCGCC	933	
DB	630	ACCACCACTCACTACCAACATCACTATCTACTACCAACCAACCAACCACTCACT	689	
QY	934	GGCGGACCAAGGTATATACCGTGGTGCGATTGACAGACGACACATGCCAGCTTC	993	
DB	690	ACCACCAACCAACCACTATCACCACCAACCAACCAACCAACCAACCACTCACTACC	749	
QY	994	TCACGAGGGGACGACGCGGACGGAAGGCTCAAGCGTCAAGCGGAAGTGTGCGCCCGGGCTT	1053	
DB	750	ACACCAACCATCACTACCACTACCACCACTATCCACCACTATGCCACCAACCAACCAACCATC	809	
QY	1054	GACATCATAGCCCGCGCCGACGGGAACCAACATGGACACCCGATAAACGACTACTAC	1113	
DB	810	ACTACGACCAACCAACCACTACTGCTGCCACCAACCAACCAACCACTCACTACCAACCA	869	
QY	1114	ACCAAGGCTCTGTGAACCAACGATGCCACCGCGACGTTTGGGGGTGGCGGCTCATC	1173	
DB	870	ACCATCAGCAACCAACCACTCACTACCAACCAACCAACCAACCACTCACTACCAACCA	929	
QY	1174	CTCCAGGCGCACCGAGCTGGACCCGAGCAAGGTGAAGACCGGCCCTATCGACCGCC	1233	
DB	930	ACCAACCACTCACTACCAACCAACCACTACCAACCAACCAACCACTCACTACCAACCA	989	
QY	1234	GACATAGTCGCCCCAGAGGAGTACGGAGCATGCGCTACGCTGGGTAGGTGAAGCTC	1293	
DB	990	ACCACCAACCACTCACTACCAACCAACCAACCAACCACTCACTACCAACCAACCAACCACT	1049	

```

RESULT 6
US-09-864-761-19241/c
: Sequence 19241, Application US/09864761
: Patent NO. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED
: TITLE OF INVENTION: GENE EXPRESSION ANAL
: FILE REFERENCE: Aequica-x-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/532,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24253.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/005666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00657

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10241
; LENGTH: 1075
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL078472.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 25
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
; OTHER INFORMATION: EST HUMAN HIT: AV739739.1, EVALUE 1.00e+00
; OTHER INFORMATION: NT HIT: AL163201.2, EVALUE 2.00e-19
;
Query Match
Best Local Similarity 4.9%; Score 97.8; DB 10; Length 1075;
Matches 452; Conservative 0; Mismatches 517; Indels 14; Gaps 3;
QY 885 GCGCCGCGGCAACAGCGGCGCGACCTACACGTCGGCTACCGCGCGCGCGGAGCAA 944
DB 1021 CACCAACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 962
QY 945 GGTCTAACCCTCGGTTCAGTTTCAGCAAGCAACCAACCAACCAACCAACCAACCA 1004
DB 961 CACCAACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 902
QY 1005 ACCGACCGCGGCAACGAGGCTCAACCGGAGTCTGCGCGCGCGCGGCTTCCAGCAG 1064
DB 901 CACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 842
QY 1065 CCGCGCGCGGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1124
DB 841 CATGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 782
QY 1125 TGGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1184
DB 781 CACCAACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 722
QY 1185 CCGGAGCTGGAGCCCGGCAAGGTCAGAGCGGCGCTTCATGAGACCGCGGACATAGTC 1244
DB 721 CACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 662

```

```

QY 1245 CCCCAGGAGATAGCGGACATCGCCTACGGTGGGGTAGGGTGAACGCTCTACAGGCCAT 1304
DB 661 CACCATCACCACCTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 602
QY 1305 CAAGTACCAACCACTAGCGCAAGCTCACCCTTCCAGCGGTCGGTCGGGACAGGAGGCC 1364
DB 601 CATCACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 542
QY 1365 CACCCACACCTTCGAGCTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1424
DB 541 CACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 483
QY 1425 GGGCTCGAGGAGATCGACCTCTACCTCTAGGACCCCAACGAGGAGGTTGACTACTC 1484
DB 482 -----TCACCATCACCACCTACCACTACCACTACCACTACCACTACCACTACCA 432
QY 1485 CTACACCCCTACTAGCGGTTTCGAGAGGTCGGCTACTACACCGGACCGCGGCGGAG 1544
DB 431 CCACGACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 374
QY 1545 GACGGTCAAGCTCGTCAAGCTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1604
DB 373 TACCACCTACCTTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 314
QY 1605 CGGAGCGCTCAGCGGAGTCCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1664
DB 313 CACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 254
QY 1665 CCGGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1724
DB 253 CACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 194
QY 1725 CACCTTCACCACTGAGGTCACACCGGTCGCAACCAACCAACCAACCAACCAACCAAC 1784
DB 193 TACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 134
QY 1785 TACTTCTACA--ACGACCTCGACCTCTACCTCTACGACCCCAACGAGGAGGTCGTTGAC 1842
DB 133 CACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 74
QY 1843 AGGTCCACGTCGAGCAACGACTA 1865
DB 73 CACACTGCACCTGATCCACCACTA 51

```

RESULT 7

```

US-09-864-761-2513/c
; Sequence 2513: Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: A60m12-1
; CURRENT APPLICATION NUMBER: US 09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30

```

[illegible]

	Query Match	4.9%	Score 97.6	DB 10:	Length 2712:
	Best Local Similarity	45.5%	Pred. No. 5,7e-16		
	Matches 509	Conservative	0	Mismatches 594	Indels 15
					Gaps 4:
QY	859	TGSGAGCGCGGTATAGTAGTCTCGTCGCGCGCGGCAACAGCGCGCGCAACACCTACAC			918
DB	1046	TGGGACTCTCCCGCGCGGAACAAAGGTGACCTCGCCTTGGGACGCGGAGCTCACCACAC			1105
QY	919	ETCGGCTTCACCGCGCGCGGCAAGGTGATACCGTTCGGTCAGTTGACAGCAGCAG			978

Db 1106 GCGACCACTGACCGCCCAAGAACAGAGCTGGCGGGAGGCTCGCCCGGGCGCTCG 1165
Qy 979 AACATCGCAGCTTCTCACAGAGGACGACCGCGGAGGAGGCTCAAGCCGGAATGC 1038
Db 1166 GTCAGCTTCGGCTTCAAGAGGACCGCGCCCGGACCCCTCGGGGTGCAAGCTCAAGGC 1225
Qy 1039 GTCGCGCCCGGGCTTGACATATAGCCCGCGCGCGGAGGAAACCAAGCATGGGACCGCG 1098
Db 1226 GCTCTCGCGAGCGCGGAGGCTCGCCCGGACACCGCCCGGCGCGCGCGCGCGCGCG 1285
Qy 1099 ATAAAGCACTACTACACCAAGGCTCTTGGAACAGCATGGCCACCCCGACGTTTCGGCG 1158
Db 1286 ACCGCGAGTGACCTCAACGAAGTCTGCTGAAGCTCTCTGGAAGCGGCCACCGACAC 1345
Qy 1159 GTTGGCGGCTCATCTCCAGCGCCACCGGAGTGGACCGCGGACAGGTTGAAGACCGCG 1218
Db 1346 AAGGCGGTCAAGAACTACGACCTCTGCGGAGCGGCGCAAGTGGCGACCGTCAACCGCG 1405
Qy 1219 CTCATCGAGACCGCCGACATAGTGGCGCCCAAGGAGATAGCGGACATCGGCTACGGTGG 1278
Db 1406 ACCACCTTCACCGACCGAAGCTTCGCCCGC---GGCAAGGACTACTCTACTCGGTCCAG 1462
Qy 1279 GGTAGGTGAAGCTCTACAAAGGCAATCAAGTACGAGCACTAGCGGACATCGACCTTACGAC 1458
Db 1463 GCGCGGACACCGCGGACACCGCGCGGTGAGCGCGCGGTCAAGGTCAACCGACCGCG 1522
Qy 1339 GCGTCCGTCGCGACAAAGGAGCGCCACCGACACCTTCGAGCTACGCGCGCGCGCGCG 1398
Db 1523 GCGGACGCGACG 1582
Qy 1399 GTGACCGCGACCTCTACTGGGACACCGGCTCGAGGACATCGACCTTACCTTACGAC 1458
Db 1583 GAGTGGGCGGTCTAGCGCGCACTACACGCTCAAAAGCTGTGACCTCCGCGTCCCGC 1642
Qy 1459 CCCAAGCGAAGAGTTGACTACTCTACACCGGCTACTACGGCTTCGAGAAGTGGCG 1518
Db 1643 GAGAAGATCACCGACATCACTACTCTCTCGCAACGTCTCAAGGCGGCAAGTGCACCATC 1702
Qy 1519 TACTAACCCGACCGCGGAACTCTGAGGTTCAAGGTCTGACCTACGAGGCGCGCGCG 1578
Db 1703 GGTGACAGCTTCGCGCGCTAGCAAGGCTTACACCGCGCGGAGTGGTTCGCGCGGTC 1762
Qy 1579 AACTACAGGTCTGCTACGAGGAGGAGCTTACGAGCTTCAAGGCGCGCGCGCGCGCG 1638
Db 1763 GCGGACACCTGGGACCGAGCGCTGCGCGGCACTTCAACAGCTTCGCAAGCTCAAGGCG 1822
Qy 1639 AATCGAAGCGAAGCGGAAACCGAAGCGGACCGGAGCTTCAAGGCTTCAAGGCTTCC 1698
Db 1823 AA---GTACCGCGACATCAAGGTCTCTGCTCTGCGCGCTTGGAGCTTGGTTCGCGCGC 1879
Qy 1699 GTTAAGGACTACTGGGACACCGAGACCTTCAACATGAAGCTCAAGCGGTGCCACCG 1758
Db 1880 TTCACCGAGCGGTGAAGAACCGCGCGCTTCCGCAAGTCTGCGAGCTTGGTTCGAG 1939
Qy 1759 AAGATAACCGGTGACCTTGCATCTTCTTACAGCACTCGACCTTCACTCTAC 1818
Db 1940 GACCGCGCTGGGCGGAGCTTTCGA-----CGGATCGACCTCGAGTGGAGTACCGG 1993
Qy 1819 GACCGCAAGCGCACTCTGTGACAGGTCCACGTCGAGCAAGCTTACGACGAGCTCGAG 1878
Db 1994 AACCGCTTCGCGCTTCAGCTCGACAGCTTCGCGTCCG---GCGCGCTCAAGACATGGTC 2050
Qy 1879 TACGCAACCGCGCGCGGAACTTGGAGCTTCTCTAGCGCTACGACCTTACGCG 1938
Db 2051 CAGCGATTCGCGCGCGGAGTTCGGACGAGCTTGGTTCACCGCGGATCACCGCGCGCGC 2110
Qy 1939 TGGGGGACTACGACTCAAGCGCGCTGTGTACTACG 1976
Db 2111 AGCTCCGCGGCAAGCTCAAGCGCGGAGTACGCGG 2148

RESULT 9

US-09-790-399-7
; Sequence 7, Application us/09790399
; Patent No. US2002038000A1
; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Tusk, Craig
; APPLICANT: Pribnow, David
; APPLICANT: Smith, Jonathan D.
; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
; FILE REFERENCE: NEX02/CI-CON2
; CURRENT APPLICATION NUMBER: US/09790,399
; PRIOR FILING DATE: 2001-02-22
; PRIOR FILING DATE: 1998-11-23
; PRIOR APPLICATION NUMBER: 07/829,461
; PRIOR APPLICATION NUMBER: 07/739,055
; PRIOR APPLICATION NUMBER: 07/561,968
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patencin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: having a 120 repeat of ACC flanked by fixed
; OTHER INFORMATION: fragments having NcoI restriction sites.
US-09-790-399-7

Query Match 4.3%; Score 84.8; DB 10; Length 390;
Best Local Similarity 52.2%; Pred. No. 4.7e-13;
Matches 188; Conservative 0; Mismatches 172; Indels 0; Gaps 0;
Qy 1247 CCAAGGAGATAGCGACATCGCTACGCTGCGGTAGGTGACGCTTACAGGCGCATCA 1306
Db 4 CCATGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 63
Qy 1307 AGTACGACGCTACGCGCAAGCTCAGCTTCACCGGCTCCGTCGCGGACAGGAGGAGCGCA 1366
Db 64 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 123
Qy 1367 CCGACACTTCGAGCTAGCGGCGCCACCTTCGTGACCGCCACCTCTACTCGGACCG 1426
Db 124 ACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 183
Qy 1427 GCTCGAGCGACATCGACTCTACTCTACGACCCCAAGGAGAGGAGTTGACTACTCT 1486
Db 184 ACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 243
Qy 1487 ACAGCGCTTACTAGCGCTTCGAGAGGTCGGCTACTACAACCCGACCGCGGAACTGGA 1546
Db 244 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 303
Qy 1547 CGTCAAGGTCTCAGCTACAGGCGCGGCGGAACTACAGGTGCGAGCTCTGAGCGAGC 1606
Db 304 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 363

RESULT 10
US-09-887-576-784
; Sequence 784, Application us/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret

; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1350.00151
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 784
; LENGTH: 1185
; TYPE: DNA
; ORGANISM: Oryza sativa
; US-09-887-576-784

Query Match 4.2%; Score 83.4; DB 10; Length 1185;
Best Local Similarity 45.3%; Pred. No. 2e-12;
Matches 363; Conservative 0; Mismatches 456; Indels 6; Gaps 2;

QY 1129 ACCAGATGCGCCACCCGACGTTTCGGCGTTCGGCGCTCATCTCCAGGCCACCCG 1188
DB 178 ACCAAGCGCCACCGTCTGACTAGAGAGATCTGTCGGGACACCTGCCGGCATGGC 237
QY 1189 AGCTGAGCCCGGACAGGTGACAGCGCCCTCATCGAGACCGCCACATATCGCCCC 1248
DB 238 TTCGTGTGCGAGCGCTCGGCGTTCGGCGTTCGGCGCTCATCTCCAGGCCACCCG 297
QY 1249 AAGGAGATAGCGGACATCGCTACGTGCGGTAGGGTGAAGCTGTACAGGCCATCAG 1308
DB 298 CAGCAGTCCCGGACATCGCGCAGGGGTGCACGGCCACTTCACAGAGCGCCCGAGG 357
QY 1309 TACGACACTACCGAAGCTCACTTCACCGGCTCGTCCGCGACAGAGGAGCCGCCAC 1368
DB 358 ATCGGCGCGGACCGGCGCCACATGTCGGCTAGCGGACCGGACGAGACCGCCGAGTG 417
QY 1369 CACACTTTCAGCTAGCGGCGGCGACCTTCGTGACGCGGCGGCTTACTTGGG---ACAG 1425
DB 418 ATGCGGCTCAGCGACGCTTCTGCCACCAAGCTCGGCGCGGCTCAGGAGTTCGCGAG 477
QY 1425 GCGTTCGAGCGACATCGACTCTTCTAGACCGCCACGGGACGAGGTGACTACTCC 1485
DB 478 AAGCGACTCGCGCTCGCTCAGCGCGGCGGAGCGGACCGAGTTCACGCTTGTAGTCTC 537
QY 1486 TACACGCGCTACTACGCTTCGAGAGGTGGGCTACTACACCGGCGGCGGCGGAGCTGG 1545
DB 538 AAGGACGCGCGCGCATGTGTCGCGGTCGCGGTCGACACGCTCTCATCTCCACCGAGAC 597
QY 1546 ACGGTTCAGGTGCTCAGCTACAGGCGCGGCGGAGTACGAGTTCGAGGTCTCAGCGAC 1605
DB 598 GAGGAGACGCTACCAACAGCAGATCGCGCGGCTCAAGGAGCAGTTCATCAAGCG 657
QY 1606 GGGAGCGCTCAGCTCAGGCGGCGGCGGACCGGATCAACCGCCACCGGACCGGACCG 1665
DB 658 GTCATCCCGACAGTACCTTCGACGAGAGACCACTTCACGCTCAACCGCTCGGCGCG 717
QY 1666 CCGACCGGACCGGACCGGCTTCACCGGTTCCGTTAAGGACTACTGGGACACGAGGAC 1725
DB 718 TTGCTCATCGGCGGCGGCGGCGGCGGCGGCTCAGCGGCGGCGGAGATCATCTGAC 777
QY 1726 ACCTTCACATGAGCTTAACACCGGTGCCAGCAGATACCGGTGACCTGACCTTCGAT 1785
DB 778 ACCTAGGGGGATGGGCGCGGCGGCGGCGGCGGCTTCGCGGCGGAGGAGCGGCGGAG 837
QY 1786 ACTTCTTACAGGCTCGACTCTTACCTCTTACGACCGCCACCGGCGGAGGCTGTGAGG 1845
DB 838 GTCGACGCGGCGGCGGCTTACATCGGCGGAGGCGGCGGAGGAGCA---TGTGCGGCGG 894
QY 1846 TCCAGTTCGAGCAGCTACGAGCAGTTCGAGTACCGGACCGGCGGCGGCGGAGCTGG 1905
DB 895 GGCCTCGCGCGGCTGCTGATGTCGAGGTGCTAGCGCATCGGCTCGCGGCGGCGGCTC 954

QY 1906 AGCTTCTCTCTTACGCTACGACGACCTACGGCTGGGGGAGCTACGAGCTCAAGCGCTC 1965
DB 955 TCGGTGTCTGTCGCTGCTTCTTACGCGGACCGGCAAGATCCCGGACAGGAGATCTCTCAAGTC 1014
QY 1966 GTCTA 1970
DB 1015 GTCAA 1019

RESULT 11
US-09-922-217-1058
; Sequence 1058, Application US/09922217
; Patent No. US200200764141
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1058
; LENGTH: 15720
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-922-217-1058

Query Match 4.2%; Score 83.2; DB 10; Length 15720;
Best Local Similarity 46.2%; Pred. No. 9.8e-12;
Matches 277; Conservative 0; Mismatches 323; Indels 0; Gaps 0;

QY 1289 AGCTCTACAGGCGCATCAAGTACGACGACTACGCCAAGCTCACCTTCACCGGCTCGTGC 1348
DB 5680 ATGACAACCCACCCACGACAGAGACCCCACTCCGCCAACACGACACCCATCACCACCC 5739
QY 1349 CGGACAGGCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1408
DB 5740 ACTAGGTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 5799
QY 1409 CCTCTACTGGGACAGGCGGCTGAGGAGATCGACCTTACTCTAGGACCGGCGGCGGCGG 1468
DB 5800 ACCACGACCTAGGTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 5859
QY 1469 ACGGTTGACTACTCTTACACCGGCTTACTAGGCTTCGAGAGGTTCGGTCTACTACAC 1528
DB 5860 ACACCATCACACCCACCTACGCTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 5919
QY 1529 CGACCGCGGAGCTGAGCGGTCAAGTCTAGCTACAAGGGCGGCGGCGGCGGCGGCGGCGG 1588
DB 5920 CCNACGACGACCCATCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCC 5979
QY 1589 TCGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1648
DB 5980 ACACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6039
QY 1649 CCNACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1708
DB 6040 CCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6099
QY 1709 ACTGGGACGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1768

Db 6100 ACCCAACACCCCGGCGACACAGACACCCCAACACCGACACCATCACCACCACTACG 6159
Oy 1769 GTGAGCTGACCTTGTGATCTCTTCTAACAGCAGCTCGACCTGTACTCTTACGACCCCAACG 1828
Db 6160 GTGAGCCGACCCCAACACCCCGGCGACACAGACCCCAACACCGACACCATCACCACCC 6219
Oy 1829 GCAACCTCTGTGACAGCTCCAGCTCGAGGACAGCTGAGGACAGCTGCGAGTACGCCAAC 1888
Db 6220 ACCACTAGCGTGACCCCAACCCCAACACCGACACCCCGGCGACACAGACCCCAACCGACCC 6279

RESULT 12

US-09-833-263-1058
; Sequence 1058, Application US/09833263
; Patent No. US2002010547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Neagher, Madeline J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 1058
; LENGTH: 15720
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-263-1058

Query Match 4.2% Score 83.2; DB 10; Length 15720;
Best Local Similarity 46.2%; Pred. No. 9.8e-12;
Matches 277; Conservative 0; Mismatches 323; Indels 0; Gaps 0;
Oy 1289 AGGTCTACAGGCGGCATCAAGTACGAGCACTACGCAAGCTACCTTACCGGCTCGGTCG 1348
Db 5680 ATGACACACACACACACAGACACCCCACTCCGCCAACCCAGCAGACCCATCACCACCC 5739
Oy 1349 CGGACAGGGAAGCGCCACACACCTTCCAGCTGACGGGGCCACCTTCTGTCAGCGCA 1408
Db 5740 ACTAGGTGACCCCAACCCCAACACCCAGCAGACCCCAACACCGACACCCCAAC 5799
Oy 1409 CCTCTACTGGGACAGGGGCTCGAGGACATGAGCTCTCTACTAGCCGCGGCGGCGG 1468
Db 5800 ACCACCACTACTAGGTGACCCCAACCCCAACACCCAGCAGACCCCAACACCC 5859
Oy 1469 AGGAGTTGACTACTCTACACCCGCTTACTAGGCTTCGAGAGGTCGGCTACTACAAC 1528
Db 5860 ACACCAATACACCAACCACTACGGTGACCCCAACCCCAACACCCAGCAGACCCCA 5919
Oy 1529 CGACCCCGGAACCTGGAGGTCAGGTCTGAGTACAGGCGGCGGCGGCACTACACG 1586
Db 5920 CCAACCAACACCACTACAGCAGCAGTACGGTGACCCCAACCCCAACACCCAGCAG 5979
Oy 1589 TCGAGCTCTGAGGACAGGGGCTTCAGCCAGTCCGGCGGCGGCGGCGGCACTACCAAC 1648
Db 5980 ACACACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAAC 6039
Oy 1649 CCAACCCGAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 1708
Db 6040 CCAACCCGAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 6099
Oy 1709 ACTGGGACACGACGACCTTACGATGACGTCACAGCGGTGCCCAACCAAGATACCG 1768
Db 6100 ACCCAACACCCCGGCGACAGACCCCAACCCCAACCCCAACCCCAACCCCAACCC 1828
Oy 1769 GTGAGCTGACCTTGTGATCTCTTCTAACAGCAGCTCGACCTGTACTCTTACGACCCCAACG 1828
Db 6160 GTGAGCCGACCCCAACACCCCGGCGACACAGACCCCAACCCCAACCCCAACCCCAAC 6219

Oy 1829 GCAACCTCTGTGACAGCTCCAGCTCGAGGACAGCTGAGGACAGCTGCGAGTACGCCAAC 1888
Db 6220 ACCACTAGCGTGACCCCAACCCCAACACCGGCGACACAGACCCCAACCGACACCC 6279

RESULT 13

US-09-974-300-1934
; Sequence 1934, Application US/09974300
; Patent No. US2002014672A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974.300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680.598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279.526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1934
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1934

Query Match 4.2% Score 82.6; DB 10; Length 1329;
Best Local Similarity 48.3%; Pred. No. 3.5e-12;
Matches 376; Conservative 0; Mismatches 349; Indels 54; Gaps 3;
Oy 515 CTCATCCAGATCTCCAGGAAAGTAATTTGGTGGGTAGATTTTGTCAATGGTAGGAGTT 574
Db 479 CTCAGAGATCTTGAAGGAGGATCAGGGCTTTTCAAGACTTTATCAACCAGAGACAG 538
Oy 575 ATCCATAGCAGCATGGACATGGAACTCATGTAGCTTCAATAGCAGCTGGTACTGGAG 634
Db 539 AACCTTATGATGACAAATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 598
Oy 635 CAGCAAGTAATGCAAGTACAGGAAATGGCTCCAGAGCTAAGCTGGCGGGAATTAAGG 694
Db 599 CGGCTCATCGGTCAGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 658
Oy 695 TCTAGGTCGGGATGTTCTGGAGCATATCTACTATAATTAAGGAGTTTCACTGGGCGG 754
Db 559 TATTGCAAAATGGGATCCGATCGCTCGAAACCGCTATTCAAGCGGTAGATTGGTGA 718
Oy 755 TTGATAACAAGATTAAGTACGGAAATTAAGGTCAATTAATCTTCTTCTTCTTCTTCT 808
Db 719 TTCAATTAAGGAAATCTGATGATCCGATCGACATTAATTCATGTCATTTGGGTG 778
Oy 809 -----GCCAGAGCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 856
Db 779 CAGAAAGCTTGGCGTACAGGAAATGAGAGAGATCCAGTCTGTTAAAGCTGTTTCA 838
Oy 857 CCGGAGCGCGGTATGATGATCTGCGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 916
Db 839 CATGGAGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 898
Oy 917 CCGTGGCTACCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 973
Db 899 CGATTGCGACCGCGGCTGTCAGCAGCAAGATTTATACAGTGGAGGCTTGGATGACAGG 958
Oy 974 -----ACGACCAATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1015
Db 959 ATACAGTCAGCGGCGGAGGATGAGTGTGCGCTCTTATTAAGCAGAGGCGGCGGCA 1018
Oy 1016 ACGGAGGCTCAGCGCGGAGTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1075
Db 1019 ATGTCAGTCAAAACCGGAGTTCTGTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1078

QY	1076	CGGGAACAGATGGGCAACCCGATAA- 	-----CGACTACTACACCAAG	1120
Db	1079	CGGATGTTTTCGATAGCTGCAAAACAAACAGAGCTCGGCACAAATATATGACAT	1138	
QY	1121	CCTCTGAACACAGATGCGCACCCGACAGTTTCGGGGTGTGGCGCTCATCTCCAGG	1180	
Db	1139	TGTCGGAGACTCGATGGCTACGCCGATCTCGGAGGAATTGCCACTTATCTCTTCAGC	1198	
QY	1181	CCACACCGAGCTGGACCCCGGCACAGGTCAGACGCGCCTCATCGAGACGCGCGACATA	1239	
Db	1199	AAGCCCGGCGCAGACAGCTGATGATGATCAACAGCTTGCTTATGGACCGGACCGATTTA	1257	

```

1 RECORD 14
2 US-09-974-300-5249/c
3 ; Sequence 5249, Application US/09974300
4 ; Patent No. US20020146721A1
5 ; GENERAL INFORMATION:
6 ; APPLICANT: Gerka, Randy M.
7 ; APPLICANT: Clausen, Ib Groth
8 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
9 ; TITLE OF INVENTION: Expression
10 ; FILE REFERENCE: 10065 500-US
11 ; CURRENT APPLICATION NUMBER: US/09/974,300
12 ; CURRENT FILING DATE: 2001-10-05
13 ; PRIOR APPLICATION NUMBER: 09/680,598
14 ; PRIOR FILING DATE: 2000-10-06
15 ; PRIOR APPLICATION NUMBER: 60/279,526
16 ; PRIOR FILING DATE: 2001-03-27
17 ; NUMBER OF SEQ ID NOS: 8481
18 ; SOFTWARE: FastSeq for Windows Version 4.0
19 ; SEQ ID NO 5249
20 ; LENGTH: 863
21 ; TYPE: DNA
22 ; ORGANISM: Bacillus clausii
23 ; FEATURE:
24 ; NAME/KEY: misc_feature
25 ; LOCATION: (1)...(863)
26 ; OTHER INFORMATION: n = A,T,C or G
27 US-09-974-300-5249

```

Query Match	4.1%	Score 80.4	DB 10	Length 863
Best Local Similarity	45.1%	Pred. No. 1.le-11		
Matches 286	Conservative	0	Mismatches 350	Indels 0
Gaps				
Qy	834	CCTCAGTCAGCGCTGACACAGCGCTGGGACCGCGTATAGTCTGCGTCGGCGCGG	893	
Db	673	CATCAGCATCGGCTGACGATCAGCGTCNGCATCGGCATCAGCTCAGCGTCNGCATCGG	614	
Qy	894	CAACAGCGGGCCGACACACCTACACCGCTCGCGCTACCCGCGCGCGACGACGGTCATAAC	953	
Db	613	CATCGCGCTCGCGCTCAGCATCCGATCGCGGCTCGGCATCGGATCGGCATCGG	554	
Qy	954	CGTCGGTCGATGTCACAGCAAGACACATCGCCAGCTTCTCCAGCAGGCGACCGACCGC	1013	
Db	553	CATCGCGCTCGCGCATCGCATCGGNTCGGATCCCGTCAGCATCAGTCGGCATCGG	494	
Qy	1014	GGACGAGAGGCTCAAGCCGCAAGTCGTCGGCCCGCGCTTGACATCATAGCCCCCGCGC	1073	
Db	493	CGTCGGCATCGGCATCGGCATCGGGCTCGGAATCCGCATCNGCATCGGCCTCGCGATCGG	434	
Qy	1074	CAGCGGACCGAGTCATCGGCACCCGACCGGATTAAGCACTACTACACCAAGCCCTCTGGACACAG	1133	
Db	433	CATCCGCATCGGCATCGGCATCAGCGTACGGCTCAGCGTCGGCATCGGCATCGGCATCGG	374	
Qy	1134	CATGGGCACCCCGCACGTTTCGGGGGTTTGGCGGCTCATCTCCAGGCCACCCGAGGTG	1193	
Db	373	CGTCGGCATCCGCATCGCGCTCAGCGTCGGCATCGGCATCGGCATCGGCATCGGCATCGG	314	
Qy	1194	GACCCCGGACAGGTTGAGACCGCCCTCATCGACACCGCGGCACATAGTGGCCCCCAAGGA	1253	
Db	313	CGTCGGCATCCCGCTCAGCATCAGCGTCGGCATCAGGCTCCGGCTCGGCATCGGCATCGG	254	

```

Db 450 CAGCTCCGGTCCGGCCGGCTGAAGAACATGTTCCAGCGATGCGCGCCCACTTCGGCAC 509
Qy 1326 GCTACCTTTACCGGCTCCGTCCGCAAGGAGGAGCCACACACCTTCGAGCTCAG 1385
Db 510 CGACTGGTTCACCGCGCCATCACCGCGAGCTCCGGCGGCAAGCTCGAGCGCC 589
Qy 1386 CGCGCCGACCTTCGTGACCGCACCTCTACTGGACACGGGCTCGAGGACATCGACCT 1445
Db 570 CGACTAGCGGGCGCCCGCCAGTACTTCGACTGGTACACGTGATGACTACGACTTCTT 629
Qy 1446 CTACCTCTACGACCCCAACGGGAACGAGTTGACTACTCTACACCGCTACTACGGCTT 1505
Db 630 CGCGGCTGGGACAGACCGCGCCGACCTCGCGCCCTGAACCTCTACAGCGG 689
Qy 1506 CGAGAGGTTCGGCTACTACACCCGACCGCGGAACTGGACGTTCAAGSTCGTCAGCTA 1565
Db 690 CATCCCAAGGCCGACTTCCACTCGCGCGCGCCATCGCCAGCTCAAGCGGAGGCG-- 747
Qy 1566 CAAGGGCGCGGCACTACGAGTCTGCTGCTCAGGAGGGAGCTCAGCCAGTCCGG 1625
Db 748 ----GTCCCGCGAGCAAGTCTCTCGGATCGGCTTCTACGGCCCGGCTGGACCG 803
Qy 1626 CGCGGGCAACCCGAATCCAAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAAC 1685
Db 804 CUTCACCGAGAGCGCCCGCGGCGCACCGCCACCGCGCGGCGCACCGCGCACCTACGAGCG 863
Qy 1686 CTTCACCGGTTCCGTTAAGCACTACTGGGACACCGAGGACACTTCACCATGACGTCAA 1745
Db 864 GGGCATCGAGGACTACAAGTCTCTAAGAACACTCTCCCGCCACCGCGCCCTCGGCGG 923
Qy 1746 CAGCGGTCCCAACAGATACCGGTGACCTTCGATAGTTCTCTACACGACCTCGA 1805
Db 924 CACCGCGTACCGCAAGTGGCGGAGCAACTGGTGGAGCTACGACACCCCGCCACCAACAA 983
Qy 1806 CCTCTACTCTACGACCCCAACGGCAACCTCTGTTGACAGTCCACGTCGAGCAACAGTA 1865
Db 984 GACCAAGATGACCTGGGCAAGGACACAGGCTCGCGCGGCGCTTCTTCGAGGTTAG 1043
Qy 1866 CGAGCAGTGGAGTACCGCAACCCCGCGGAGACCTGGAGCTTCCTGCTCTA 1919
Db 1044 CGGTGACACGGCGGCGGAACCTGGTCTCCGGATGAGACTCCGGGCTCCGCTA 1097

```

Search completed: November 2, 2002, 13:09:07
 Job time : 89.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 2, 2002, 08:16:03 ; Search time 51.5 seconds
(without alignments)
11772.815 Million cell updates/sec

Title: US-09-841-553-6

Sequence: 1 ATGAAGGGCTGAAGCTCT.....AGGCCCTCGTCTACTACGGG 1977

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptcdat1/1/lna/5A_COMB.seq:*
- 2: /cgn2_6/ptcdat1/1/lna/5B_COMB.seq:*
- 3: /cgn2_6/ptcdat1/1/lna/6A_COMB.seq:*
- 4: /cgn2_6/ptcdat1/1/lna/6B_COMB.seq:*
- 5: /cgn2_6/ptcdat1/1/lna/6C_COMB.seq:*
- 6: /cgn2_6/ptcdat1/1/lna/6D_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1977	100.0	1977	4	US-08-894-818B-6
2	1366.4	69.1	1962	4	US-08-894-818B-34
3	1366.4	69.1	1962	4	US-09-445-472-15
4	1360.8	68.8	1977	4	US-08-894-818B-2
5	1360.8	68.8	1977	4	US-08-445-472-11
6	1165	58.9	1859	4	US-08-894-818B-15
7	969.4	49.0	1356	4	US-08-894-818B-4
8	837	42.3	1236	4	US-09-445-472-2
9	669.8	33.9	898	1	US-08-750-532-7
10	507	25.6	507	4	US-08-894-818B-26
11	427	21.6	564	1	US-08-750-532-11
12	427	21.6	564	4	US-08-894-818B-14
13	341.6	17.3	419	4	US-08-894-818B-30
14	197	10.0	1464	4	US-08-894-818B-17
15	131	6.6	2539	3	US-09-000-016-3
16	131	6.6	2539	4	US-09-514-340-3
17	131	6.6	2809	3	US-09-000-016-1
18	131	6.6	2809	4	US-09-514-340-1
19	107.6	5.4	5163	3	US-08-700-651-1
20	107.6	5.4	5163	3	US-08-928-361B-4
21	107.6	5.4	5163	3	US-08-700-651-2
22	107.6	5.4	5318	3	US-08-928-361B-3
23	98	5.0	4403765	4	US-09-103-840A-2
24	97.6	4.9	2712	3	US-09-025-591-4
25	97.4	4.9	5511	3	US-08-928-361B-2
26	97.4	4.9	7334	3	US-08-928-361B-1
27	86.8	4.4	2064	1	US-08-343-428-1

28	84.8	4.3	390	4	US-09-197-649-7	Sequence 7, Appl
29	82.2	4.2	1140	3	US-09-023-173-4	Sequence 4, Appl
30	76.6	3.9	2793	1	US-08-309-747-1	Sequence 1, Appl
31	76.6	3.9	2793	1	US-08-458-299-1	Sequence 1, Appl
32	76	3.8	12588	2	US-08-387-942C-1	Sequence 1, Appl
33	75.8	3.8	3468	1	US-07-951-715A-2	Sequence 2, Appl
34	75.8	3.8	3468	2	US-08-459-504B-2	Sequence 2, Appl
35	75.8	3.8	3468	3	US-08-459-504B-2	Sequence 2, Appl
36	75.8	3.8	3468	3	US-08-459-504B-2	Sequence 2, Appl
37	75.8	3.8	3468	3	US-08-459-504B-2	Sequence 2, Appl
38	75.8	3.8	3468	3	US-08-459-504B-2	Sequence 2, Appl
39	75.8	3.8	3468	3	US-08-459-504B-2	Sequence 2, Appl
40	74.2	3.8	1209	4	US-09-547-432-2	Sequence 3, Appl
41	74.2	3.8	1513	1	US-08-314-309A-5	Sequence 2, Appl
42	74.2	3.8	1513	1	US-08-314-309A-2	Sequence 2, Appl
43	73.2	3.7	2753	2	US-07-952-853-5	Sequence 5, Appl
44	73.2	3.7	2753	2	US-08-914-848-5	Sequence 5, Appl
45	72.8	3.7	2185	1	US-08-173-508-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-894-818B-6
; Sequence 6, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, HiKaru
; APPLICANT: MORISHITA, HIO
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: NITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Saburo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 33285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA

Thu Nov 7 10:12:17 2002

US-08-894-818B-6

Query Match		100.0%	Score 1977:	DB 4:	Length 1977:	O:		Gaps	
Best Local Similarity		100.0%	Pred. No. 0:	Mismatches		Indels		0:	
Matches 1977:		Conservative	0:	0:		0:		0:	
Qy	1	ATGAGGGCTCAAGCTCTCAVATAGTATGATTTAGTCTAGGTTGGTGTAGTAGGAGC	60						
Db	1	ATGAGGGCTCAAGCTCTCAVATAGTATGATTTAGTCTAGGTTGGTGTAGTAGGAGC	60						
Qy	61	GTAGCGAGCTCCAGACAGAAAGTGTACAGTAAAGATGTTGAGAAAGTATAGT	120						
Db	61	GTAGCGAGCTCCAGACAGAAAGTGTACAGTAAAGATGTTGAGAAAGTATAGT	120						
Qy	121	CTGCTAACGCGAGGACTGTTCCAGAAAATTCAAAATTTGAATCCTAACGAGAAATCAGC	180						
Db	121	CTGCTAACGCGAGGACTGTTCCAGAAAATTCAAAATTTGAATCCTAACGAGAAATCAGC	180						
Qy	181	ACAGTAATCTATTTGAAACCATAGGAAAGAAATGTCAGTAAAGTCTTTCAGTTA	240						
Db	181	ACAGTAATCTATTTGAAACCATAGGAAAGAAATGTCAGTAAAGTCTTTCAGTTA	240						
Qy	241	ATGGGTGCAAAAGTATAGTCTGATACCATATTTATACCGCAATAGTCCGATCTTAAG	300						
Db	241	ATGGGTGCAAAAGTATAGTCTGATACCATATTTATACCGCAATAGTCCGATCTTAAG	300						
Qy	301	GTAGAGACTTACTAGTATCTCAGGTTTACAGGGGGTAAAGTAAAGTCTTTCAGGTGT	360						
Db	301	GTAGAGACTTACTAGTATCTCAGGTTTACAGGGGGTAAAGTAAAGTCTTTCAGGTGT	360						
Qy	361	AGGTATATCCAGGAAGTACAAAGTTTACAGTTTACAGAGAAATTAAGAGACTTGAATG	420						
Db	361	AGGTATATCCAGGAAGTACAAAGTTTACAGTTTACAGAGAAATTAAGAGACTTGAATG	420						
Qy	421	TCGCAAGCTCAAGTATGCAACTTACGTTTGGAACTTGGGATATGATGTTCTGGAATC	480						
Db	421	TCGCAAGCTCAAGTATGCAACTTACGTTTGGAACTTGGGATATGATGTTCTGGAATC	480						
Qy	481	ACATAGGAATTAATGACATGGAATTCAGGCTTCATCCAGATCTCCAAGGAAGTA	540						
Db	481	ACATAGGAATTAATGACATGGAATTCAGGCTTCATCCAGATCTCCAAGGAAGTA	540						
Qy	541	ATTGGGTGGTATGTTTGTCAATGGTAGGATTCATCATACATACATGACCATGGACATGGA	600						
Db	541	ATTGGGTGGTATGTTTGTCAATGGTAGGATTCATCATACATACATGACCATGGACATGGA	600						
Qy	601	ACTCATGTAGCTTCATAGCAGCTGGTACTGGAGCAGCAAGTAAATGGCAAGTACAAAGGA	660						
Db	601	ACTCATGTAGCTTCATAGCAGCTGGTACTGGAGCAGCAAGTAAATGGCAAGTACAAAGGA	660						
Qy	661	ATGGCTCCAGGAGCTAAAGCTGGGGGANTTAAGTCTAGTGGCGGATGTTCTGGAAGC	720						
Db	661	ATGGCTCCAGGAGCTAAAGCTGGGGGANTTAAGTCTAGTGGCGGATGTTCTGGAAGC	720						
Qy	721	ATATCTACTAATTAAGGAGTGTAGTGGCGGCTGTATACAAAGTAAAGTACCGAAT	780						
Db	721	ATATCTACTAATTAAGGAGTGTAGTGGCGGCTGTATACAAAGTAAAGTACCGAAT	780						
Qy	781	AGGTCAATTAATCTTCTTGTTCAGCCAGAGCTCCGAGCAAGTCCCTCAGT	840						
Db	781	AGGTCAATTAATCTTCTTGTTCAGCCAGAGCTCCGAGCAAGTCCCTCAGT	840						
Qy	841	CAGGCGGTCAACACGCTGGAGCGGCTATAGTATGCTGCTGCGCGCGGCAACAC	900						
Db	841	CAGGCGGTCAACACGCTGGAGCGGCTATAGTATGCTGCTGCGCGCGGCAACAC	900						
Qy	901	GGGCGGAACCTTACACCGTTCGGCTACCGCGCGCGCGGAGCAAGTCAATACCGTCGT	960						
Db	901	GGGCGGAACCTTACACCGTTCGGCTACCGCGCGCGCGGAGCAAGTCAATACCGTCGT	960						
Qy	961	CGAGTTCAGCAGCAGCAACATCCGCAAGTTCCTCAGAGGGGAGCAGCGCGGAGGA	1020						
Db	961	CGAGTTCAGCAGCAGCAACATCCGCAAGTTCCTCAGAGGGGAGCAGCGCGGAGGA	1020						

RESULT 2

US-08-894-818B-34

: Sequence 34, Application US/08894818B

: Patent No. 6261822

: GENERAL INFORMATION:

: APPLICANT: TAKAKURA, Hikaru

: APPLICANT: MORISHITA, Mio

Qy	1021	AGCTCAAGCCGGAAGTCTGCGCCCGGGGTTGACATATAGCCCGCGCGGCGGAG	1080
Db	1021	AGCTCAAGCCGGAAGTCTGCGCCCGGGGTTGACATATAGCCCGCGCGGCGGAG	1080
Qy	1081	ACCAGCATGGGACCGCCGATTAACAGCTACTACACACAGGCTCTGGAACACGATGGCC	1140
Db	1081	ACCAGCATGGGACCGCCGATTAACAGCTACTACACACAGGCTCTGGAACACGATGGCC	1140
Qy	1141	ACCCGCGAGTTCGGCGGTTGGCGGCTCATCTCCAGGCCACCCGAGCTGAGCCCG	1200
Db	1141	ACCCGCGAGTTCGGCGGTTGGCGGCTCATCTCCAGGCCACCCGAGCTGAGCCCG	1200
Qy	1201	GACAGCTAGAGCGCCCTCATCGAGACCGCCGACATAGTCCGCCCCCAAGGAGATAGG	1260
Db	1201	GACAGCTAGAGCGCCCTCATCGAGACCGCCGACATAGTCCGCCCCCAAGGAGATAGG	1260
Qy	1261	GACATCGCTACGGTGGGTTAGGTTGAACGTTTACAGGCCATCAAGTAGGAGACTAC	1320
Db	1261	GACATCGCTACGGTGGGTTAGGTTGAACGTTTACAGGCCATCAAGTAGGAGACTAC	1320
Qy	1321	GCCAAAGCTCACCTTCACCGGCTCCGTCGCCGACAAAGGAGAGCGCACACCTTCGAC	1380
Db	1321	GCCAAAGCTCACCTTCACCGGCTCCGTCGCCGACAAAGGAGAGCGCACACCTTCGAC	1380
Qy	1381	GTGAGCGGCGCACCTTCGTGACGCGCACCCCTTACTGGGACAGCGGCTCGAGCGACATC	1440
Db	1381	GTGAGCGGCGCACCTTCGTGACGCGCACCCCTTACTGGGACAGCGGCTCGAGCGACATC	1440
Qy	1441	GACCTCTACTCTAGACCCCAAGCGGAGAGGTTGACTCTCTACACCGGCTACTTAC	1500
Db	1441	GACCTCTACTCTAGACCCCAAGCGGAGAGGTTGACTCTCTACACCGGCTACTTAC	1500
Qy	1501	GGCTTCGAGAGGTCGGTCTATCAACCCGACCGCGGAACTCGAGCGTCAAGGTCGTC	1560
Db	1501	GGCTTCGAGAGGTCGGTCTATCAACCCGACCGCGGAACTCGAGCGTCAAGGTCGTC	1560
Qy	1561	AGCTACAGGCGCGGCAACTTACAGCTGCGCTGACGAGCGGAGGCTTCAGCCAG	1620
Db	1561	AGCTACAGGCGCGGCAACTTACAGCTGCGCTGACGAGCGGAGGCTTCAGCCAG	1620
Qy	1621	TCGCGGCGGCAACCGGATCCAAACCCCAACCCGCAACCCGCAACCCGAGCAGCAGC	1680
Db	1621	TCGCGGCGGCAACCGGATCCAAACCCCAACCCGCAACCCGAGCAGCAGCAGC	1680
Qy	1681	CAGAGCTTCACCGGTTCCGTTAACGACTACTGAGGACACGAGGACCCCTTACCATGAC	1740
Db	1681	CAGAGCTTCACCGGTTCCGTTAACGACTACTGAGGACACGAGGACCCCTTACCATGAC	1740
Qy	1741	GTCAACAGCGGTGCCACCAAGATAACCGGTGACCTTCGATCTTCCTACACGAC	1800
Db	1741	GTCAACAGCGGTGCCACCAAGATAACCGGTGACCTTCGATCTTCCTACACGAC	1800
Qy	1801	CTGAGCTCTACCTCTAGGACCCCAAGCGCAACCTCGTTGACAGGTCCAGCTCGAGAAC	1860
Db	1801	CTGAGCTCTACCTCTAGGACCCCAAGCGCAACCTCGTTGACAGGTCCAGCTCGAGAAC	1860
Qy	1861	AGCTAGCAGCAGCTGAGTACGCCAACCCCGCGGGAACCTTCGCTCTCTCTAC	1920
Db	1861	AGCTAGCAGCAGCTGAGTACGCCAACCCCGCGGGAACCTTCGCTCTCTCTAC	1920
Qy	1921	GCCTACAGCACTACGGCTGGGGGACTACACAGCTCAAGGCGGCTCTCTACTACGGG	1977
Db	1921	GCCTACAGCACTACGGCTGGGGGACTACACAGCTCAAGGCGGCTCTCTACTACGGG	1977

APPLICANT: YAMAMOTO, Katsuhiko
 APPLICANT: KITTA, Masanori
 APPLICANT: ASADA, Kiyozo
 APPLICANT: TSUNASAWA, Susumu
 APPLICANT: KATO, Ikunoshin
 TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 NUMBER OF SEQUENCES: 42
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Browdy and Neimark
 STREET: 419 Seventh Street N.W., Ste. 300
 CITY: Washington
 STATE: D.C.
 COUNTRY: United States of America
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/894,818B
 FILING DATE: 20-MAY-1998
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/JP96/03253
 FILING DATE: 07-NOV-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 332385/1995
 FILING DATE: 12-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Browdy, Roger L.
 REGISTRATION NUMBER: 25,618
 REFERENCE/DOCKET NUMBER: TAKAKURA-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 528-5197
 TELEFAX: (202) 737-3528
 INFORMATION FOR SEQ ID NO: 34:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1962 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: genomic DNA
 US-08-894-818B-34

Query Match	69.1%	Score 1366.4;	DB 4;	Length 1362;
Best Local Similarity	81.2%;	Prod. No. 9.7e-232;		
Matches 1605;	Conservative 0;	Mismatches 356;	Indels 15;	Gaps
<hr/>				
Qy	1	ATGAAGGGGCTGAACGCTCTCATATTAGTGTATTAGTTCTTAGTTTGTTAGTAGTAGGAGC	60	
Db	1	ATGAAGGGGCTGAACGCTCTCATATTAGTGTATTAGTTCTTAGTTTGTTAGTAGTAGGAGC	60	
<hr/>				
Qy	61	GTAGGGCGACTCCAGAGAAGAAAGTTTGAACAAGTAACAANAATTGCAGAAGAACACTATTCCT	120	
Db	61	GTAGGGCGACTCCAGAGAAGAAAGTTGTACAGTTAAGAAATGTTGAGAAGAACACTATTCGT	120	
<hr/>				
Qy	121	CTGCTCAACGCCAGGACTCTTCAGAAAAATTCAAAANAATTGAATCCTAACGAGCAATAACGC	180	
Db	121	CTGCTCAACGCCAGGACTGTTCAGAAAAATTCAAAANAATTGAATCCTAACGAGCAATAACGC	180	
<hr/>				
Qy	181	ACAGTAATTTGATTTTGAAACCATPAGGAAAAAGAAATTCSCATPAGAGTCTTTGAGTTA	240	
Db	181	ACAGTAATTTGATTTTGAAACCATPAGGAAAAAGAAATTCGATPAGAGTCTTTGAGTTA	240	
<hr/>				
Qy	241	ATGGGTGCCAAAGTTAGTATGTGTACCATATTATACCCGCAATAGCTGCCCATCTTTAAG	300	
Db	241	ATGGGTGCCAAAGTTAGTATGTGTACCATATTATACCCGCAATAGCTGCCGATCTTTAG	300	
<hr/>				
Qy	301	CTTACAGACTTACTAGTCACTCCAGGTTTAAACAGGGGGTAAAGCTTAAGCTTTCAGGTGTT	360	
Db	301	CTTACAGACTTACTAGTCACTCCAGGTTTAAACAGGGGGTAAAGCTTAAGCTTTCAGGTGTT	360	

QY	361	AGTTTATTCAGGAAGACTACAAAGTTTACAGTTTACAGCAAAATTTAGAAAGCTGGATGAG	420
DB	361	AGTTTATTCAGGAAGACTACAAAGTTTACAGTTTACAGCAAAATTTAGAAAGCTGGATGAG	420
QY	421	TCTCGACGTCAAAGTTATGGCAACTTTACGTTTGGAACTTGGGATATGATCGTTCTGGAAATC	480
DB	421	TCTCGACGTCAAAGTTATGGCAACTTTACGTTTGGAACTTGGGATATGATCGTTCTGGAAATC	480
QY	481	ACAAATAGGAATAATTTGACACTGGAAATTCAGCGCTTCTCATCCAGATCTCCACGAAAAGTA	540
DB	481	ACAAATAGGAATAATTTGACACTGGAAATTCAGCGCTTCTCATCCAGATCTCCACGAAAAGTA	540
QY	541	ATTGGGTGGGTAGATTTTGTCAATGGGTAGAGTTATCCATACGATGACCATGGACATGGA	600
DB	541	ATTGGGTGGGTAGATTTTGTCAATGGGTAGAGTTATCCATACGATGACCATGGACATGGA	600
QY	601	ACTCATGTAGCTTCAATAGCACTGGTACTGGAGCAGCAAGTAATGGCAAGCTACMAGGA	660
DB	601	ACTCATGTAGCTTCAATAGCACTGGTACTGGAGCAGCAAGTAATGGCAAGCTACMAGGA	660
QY	661	ATGGCTCCAGCAGCTAAGCTGGCGGGAAATTAAGTCTTCTAGTCCGATGTTCTTGGAAAC	720
DB	661	ATGGCTCCAGCAGCTAAGCTGGCGGGAAATTAAGTCTTCTAGTCCGATGTTCTTGGAAAC	720
QY	721	ATATCTACTATTAATTAAGGGAGTTGAGTTGGCCGTTGATAACAAAGATAAGTACGGAAAT	780
DB	721	ATATCTACTATTAATTAAGGGAGTTGAGTTGGCCGTTGATAACAAAGATAAGTACGGAAAT	780
QY	781	AGGTCATTAACTTTCTTCTTGGTTCAAGCCAGAGACTCCGAGCAACGACTCCCTCACT	840
DB	781	AGGTCATTAACTTTCTTCTTGGTTCAAGCCAGAGACTCCGAGCAACGACTCCCTCACT	840
QY	841	CAGGCGCTCAACAAGCCTGGGACGGCGGTATCTAGTCTGGCTCGCCGCGGCACACAGC	900
DB	841	CAGGCTGTAAATGCAGCGTGGGATCTGGATTTAGTTTGTGGTTGGCGCTTGGAAACAGT	900
QY	901	GGCGCGACACCTTACACGCTCGGCTCACCCGCGCCGCGGAGCAAGGTGATAACCGTCTG	960
DB	901	GGACCTAACAGTATACAAATCCGTTCTCGCAGCGCTGCAAGGAAAGATTATTTACAGTT	960
QY	961	GCAGTTGACAGCAGCAGCAACNTCCGCCAGCTCTTCCAGCAGGGGACCGCGCGGACGGA	1020
DB	961	GGCGTGTACAAAGTATGATGTTTATTAACAGCTCTCAAGCAGAGGGGCCAACTGCAGAC	1020
QY	1021	AGGCTCAAGCGCGGAAGTCTGGCCCGCGCGTGCAGATCATAGCCCCGCGCCGACGCGGA	1080
DB	1021	AGGCTTAAAGCCTGAGTTGTTGCTCCAGGAACCTGATTAATTTGCTGCCAGACGCAAGT	1080
QY	1081	ACCAGATATGGGCAACCCGATTAACGACTACTACACCAAGGCGCTTGGACCGACATGCCC	1140
DB	1081	ACTAGCATGGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCA	1140
QY	1141	ACCCCGCAGCTTTCGGGCGTTGGCGCGTCTCATCTCCAGGCCCCACCGAGCTGGACCCG	1200
DB	1141	ACTCTCTACGTTAGCTGGTATTCGACGCCCTTCTGCTCCAGCAGACCGGAGCTGACTCCA	1200
QY	1201	GACAAGGTGACAGCCGCCCTCATCCAGACCGCGACATAGTGCGCCCCCAAGGAGATACGC	1260
DB	1201	GACAAAGTAAAAACAGCCCTCATAGAAACTGCTGATATCTGTTAAAGCAAGTGAATAGCC	1260
QY	1261	CAGATTCGCTACGTCGGGTGAGGTGAAAGTCTACAGGGCATCATGTACGACGACTAC	1320
DB	1261	GATATAGCCTACGGTGCAGGTGAGGTTAAATGTCATACAAAGCTATTAACCTACGTAAC	1320
QY	1321	GCNAAGTCACTTTCACCGGCTCCGTCGGCGACAGGGAAGCGCCACCCACACTTCGAC	1380
DB	1321	GCAAGCTAGTGTCTACTGGATATGTTGCCAACAAAGGACGCCAAACTCACCAGTTCGTT	1380
QY	1381	GTACGGCGGCGCACTTCGTGTGACCGCACCCCTCTACTCGGACACGGGCTCGAGGACATC	1440
DB	1381	ATTAGGGAGCTTCGTTCTGACTGCCACATATATCTGGGACAAATGCCAAATAGGACCTT	1440
QY	1441	GACCTCTACCTCTACGACCCCAAGCGGAGCGAGTTGACTACTCTCTACACGGCGCTACTAC	1500

841 CCCGACCCACCCGCGACCCAGACCCAGACCTTCCGCGTTACGACTACTG 900
1713 GCACACGAGGACACCTTCCACCTGACGCTCAACAGCGGTCACACCAAGATACCGGTGA 1772
901 GCACACGAGGACACCTTCCACCTGACGCTCAACAGCGGTCACACCAAGATACCGGTGA 960
1773 CCGTACCTTGGATACCTTCTACACGACCTCGACCTCTACCTCTACGACCCCAACGCA 1832
961 CCGTACCTTGGATACCTTCTACACGACCTCGACCTCTACCTCTACGACCCCAACGCA 1020
1833 CCGTACCTTGGATACCTTCTACACGACCTCGACCTCTACCTCTACGACCCCAACGCA 1892
1021 CCGTACCTTGGATACCTTCTACACGACCTCGACCTCTACCTCTACGACCCCAACGCA 1080
1893 CCGGGAACCTTGGAGCTTCTCTGCTACGCTACGACCTACGCTTGGGCGGACTACCA 1952
1081 CCGGGAACCTTGGAGCTTCTCTGCTACGCTACGACCTACGCTTGGGCGGACTACCA 1140
1953 GCTCAAGCCCGCTGCTACTACGGG 1977
1141 GCTCAAGCCCGCTGCTACTACGGG 1165

RESULT 7
US-08-894-8188-4
Sequence 4, Application US/088948188
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, HIKARU
APPLICANT: MORISHITA, MIO
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 415 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,8188
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1565 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA

TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1859 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-8188-15

Query Match 58.9%; Score 1165; DB 4; Length 1859;
Best Local Similarity 100.0%; Pred. No. 2e-247;
Matches 1165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 813 GAGCTCCGACGAGACCCGCTCCTCAGTCAGGCGCTCAACAGCGCTGGGACGCGGTAT 872
DB 1 GAGCTCCGACGAGACCCGCTCCTCAGTCAGGCGCTCAACAGCGCTGGGACGCGGTAT 60
QY 873 AGTAGTCTGCGTCCGCGCGGCAACAGCGGCGCAACACCTACGCTCGGCTCACCCTG 932
DB 61 AGTAGTCTGCGTCCGCGCGGCAACAGCGGCGCAACACCTACGCTCGGCTCACCCTG 120
QY 933 CGCGCGGACGAGCTCAATACCGTCCGCTGCAAGTTGACAGCAACGACATCGCCAGCTT 992
DB 121 CGCGCGGACGAGCTCAATACCGTCCGCTGCAAGTTGACAGCAACGACATCGCCAGCTT 180
QY 993 CTCACAGGAGGACCGCGGACGAGGCTCAAGCGAAGTCTCGGCGCGCGCGT 1052
DB 181 CTCACAGGAGGACCGCGGACGAGGCTCAAGCGAAGTCTCGGCGCGCGCGT 240
QY 1053 TGACATCATAGCCCGCGCGGCAACAGGACGAGTGGGACCGGCTAAAGCTACTA 1112
DB 241 TGACATCATAGCCCGCGCGGCAACAGGACGAGTGGGACCGGCTAAAGCTACTA 300
QY 1113 CACCAAGGCTCTGGAACAGCATGCGCCACCGCGCAAGTTGCGGGGCTTGGCGGCTAT 1172
DB 301 CACCAAGGCTCTGGAACAGCATGCGCCACCGCGCAAGTTGCGGGGCTTGGCGGCTAT 360
QY 1173 CTCACAGGAGGACCGCGGACGAGGCTCAAGCGAAGTCTCGGCGCGCGT 1232
DB 361 CTCACAGGAGGACCGCGGACGAGGCTCAAGCGAAGTCTCGGCGCGCGT 420
QY 1233 CGACATAGTGGCGGCGGCAAGGATAGCGGACATCGGCTACGCTGCGGGTGGTCAAGCT 1292
DB 421 CGACATAGTGGCGGCGGCAAGGATAGCGGACATCGGCTACGCTGCGGGTGGTCAAGCT 480
QY 1293 CTACAAGGCGCATCAAGTACGAGCTACGCGCAAGCTCCTTACGCGGCTCGGCGCA 1352
DB 481 CTACAAGGCGCATCAAGTACGAGCTACGCGCAAGCTCCTTACGCGGCTCGGCGCA 540
QY 1353 CAGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1412
DB 541 CAGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
QY 1413 CTACTGGGACACGGGCTCGAGGACATCGGACCTTACCTTACGACCCCAACGGAAGA 1472
DB 601 CTACTGGGACACGGGCTCGAGGACATCGGACCTTACCTTACGACCCCAACGGAAGA 660
QY 1473 GGTGTGACTTCTCTACACCGGCTTACGCGCTTCGAGAGGTCGGCTACTACAAACCGAC 1532
DB 661 GGTGTGACTTCTCTACACCGGCTTACGCGCTTCGAGAGGTCGGCTACTACAAACCGAC 720
QY 1533 CGCGGGAACCTTGGACGCTCAAGTCTGACGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1592
DB 721 CGCGGGAACCTTGGACGCTCAAGTCTGACGAGGCGGCGGCGGCGGCGGCGGCGGCGG 780
QY 1593 CGTCTGACGAGCGGAGGCTCAGCAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1652
DB 781 CGTCTGACGAGCGGAGGCTCAGCAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
QY 1653 CCGGGAACCCCGGACGAGGCTCAGCAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1712

FEATURE:
OTHER INFORMATION: /note= N at position 1383 is G or T.
us-08-894-818B-4

Query Match 49.0%; Score 969.4; DB 4; Length 1566;
Best Local Similarity 76.5%; Pred. No. 28-204;
Matches 1208; Conservative 0; Mismatches 357; Indels 15; Gaps 1;

QY 397 GCAGATTAGAGAGCTGGATGATGCTGCAGCTCAAGTTATGCAACTTACGTTTGGAA 456
DB 1 GCAGATTAGAGAGCTGGATGATGCTGCAGCTCAAGTTATGCAACTTACGTTTGGAA 60

QY 457 TTGGATATGATGCTTCTGGAATCAATAGGAATAATTGACACTGGAATTGAGCTTCT 515
DB 61 TTGGATATGATGCTTCTGGAATCAATAGGAATAATTGACACTGGAATTGAGCTTCT 120

QY 517 CATCCAGATCTCCAAAGAAAGTAATTCGGTGGGTAGATTTCATATGGTAGAGTTAT 576
DB 121 CATCCAGATCTCCAAAGAAAGTAATTCGGTGGGTAGATTTCATATGGTAGAGTTAT 180

QY 577 CCATACGATGACCATGGATGATGCTGATCTATGCTCAATAGCAGCTGGTACTGGAGCA 636
DB 181 CCATACGATGACCATGGATGATGCTGATCTATGCTCAATAGCAGCTGGTACTGGAGCA 240

QY 637 GCAAGTAAATGGCAAGTACAGGGAATGCTCCAGGAGCTAAGCTGGCGGGAATTAAGTT 696
DB 241 GCAAGTAAATGGCAAGTACAGGGAATGCTCCAGGAGCTAAGCTGGCGGGAATTAAGTT 300

QY 697 CTAGTGGCATGCTTCTGGAGGATATCTACTATTAATTAAGGAGTTAGTGGGCCCTT 756
DB 301 CTAGTGGCATGCTTCTGGAGGATATCTACTATTAATTAAGGAGTTAGTGGGCCCTT 360

QY 757 GATAACAAGATTAAGTACGGAATTAAGTCAATTAATTCCTCTCTCTCTCTCTCTCTCT 816
DB 361 GATAACAAGATTAAGTACGGAATTAAGTCAATTAATTCCTCTCTCTCTCTCTCTCTCT 420

QY 817 TCGAGCGAAGCGATCTCCCTCAGTCAGGCGGTCACACAGCGCTGGGAGCGGATTAAGTT 875
DB 421 TCGAGTGGTACTGAGCGCTCAAGTCAGGCTGTAATGAGCGGTGGATGCTGGATTAAGTT 480

QY 877 GTCTGCTGCTGCGCGGCAACAGCGGCGCAACAGCTTACAGCAACAGCACTCTCTCT 936
DB 481 GTCTGCTGCTGCGCGTGGAAACAGTGGACCTACACAGTATACATCTGGTTCTCCACAGCT 540

QY 937 GCGAGCAAGTCAATACCGCTCGCTGAGTTGACAGCAACAGCACTCTCTCTCTCTCTCT 996
DB 541 GCAAGCAAGTTATTAAGTTGAGCGCTTGACAAAGTATGTTATTAACAAGCTTCTCA 600

QY 997 ACAGGGGACGACGCGGACGGAAGGCTCAAGCGGGAAGTCTGCGCGCGCGGCTTGAC 1056
DB 601 ACAGAGGCGCACTGACAGCGGCGAGGCTTAAGCTGAGTTGCTCTCCAGGAAGTGG 660

QY 1057 ATCATAGCGCGCGCGGCAACAGCACTGGGCAACCGCGGATTAAGCACTACTACACC 1116
DB 661 ATAAATGCTGCGCAGCAAGTGGAACTGAGTGGTCAACCAATTAATGACTATTACAGA 720

QY 1117 AAGGCTCTGCAACCACTGCGCCAGCGGACGTTTGGGCGTGGGCGCTCTCTCTCTCT 1176
DB 721 GAGCTCTGGACATCAATGGCAACTCTCAGCTAGCTGATTAAGCGCCTCTCTCTCTCT 780

QY 1177 CAGCGCCACCGAGCTGGACCGGCAAGGTGAAGACCGCTCTATGAGACCGCGGAC 1236
DB 781 CAAGCACACCGAGCTGGATCTCAGACAAAGTAAACAGCCCTCTAAGAACTGCTGTAT 840

QY 1237 ATAGTGGCGCCCAAGGAGATAGCGGACATCGCTTACGCTGGGTTAGGTTGAAGTGTAT 1296
DB 841 ATCTGTAAGCCAGATGAATAGCCGATATAGCTTACGCTGAGGTTAGGTTAATGCTATAC 900

QY 1297 AAGGCGCATCAAGTACGAGCACTACGCCAAGTCACTCTCACCGGCTCGCTCGCGCAAG 1356
DB 901 AAGGCTATAAATCAATCAATATGCAAGCTAGTGTCTCACTGGATATGTTGGCAACA 960

QY 1357 GGAAGCGCCACCCACACCTTCGACGTGACGGGCGGCGGACCTTCTGACCGCGGCACTCTAC 1416

DB 961 GCGAGCAAACTCACCAGTTGTTATACGGGAGCTTCTCTGCTTACGTCGCAATATATAC 1020
QY 1417 TGGGACACGGGCTCGAGCGATCGACCTCTACCTCTACGACCCCAACGGGAACGAGGTT 1476
DB 1021 TGGGACAAATGCAATAGCGACCTTGTATCTTACCTCTACGATCCCAANTGGAACGAGTT 1080

QY 1477 GACTACTCTTACACGGCTTACTACGGCTTTCGAGAAGGTTCGGCTACTACAAACCGACGCC 1536
DB 1081 GACTACTCTTACACGGCTTACTATGATTCGAAAAGTTGGTTATTTACAAACCACTGAT 1140

QY 1537 GGAACCTGGAGCGGTCAAGGTCGTCACTACAGGCGCGCGGAACTACCAAGTCTGACGTC 1596
DB 1141 GGAACATGGACAATAAAGTTTCTAGCTACAGCGAAGTCAAACTATCAAGTAGATGTG 1200

QY 1597 GTACGGAGCGGAGCTCAGCGAGTCCGCGCGGCGGCAACCCGAATCCAAACCCGACCG 1656
DB 1201 GTAACTGATGTTCCCTTTCACAGGCTGGAAG-----TTTCACCATCTCCA 1245

QY 1657 ACCCAACCCCGACCCAGCACCCAGACCTTTCACCGCTTTCGCTTAAAGACTACTGGGAC 1716
DB 1246 CAACCAAGAACCAACAGCTAGACGCAAGAGCTTCCMAGNATCGSATCACTACTATGAC 1305

QY 1717 ACCAGCAGACCTTCAACATGAACGTCAACAGGCTGCCACCAAGATTAACCGGTGACCTG 1776
DB 1306 AGGAGCAGACCTTCAACATGAACGTTAACTCTGGGCTTACAAAGATTAAGTGGAGACCTA 1365

QY 1777 ACCTTCGATCTTCTTACAGCAACGCTCGACCTCTACTCTAGCACCCCAACGCAACCTC 1836
DB 1366 GTGTTGACACAGCTTACCTGATCTTGAACCTTACCTCTAGCATCTTACCAAGACTT 1425

QY 1837 GTTGACAGCTTCCAGCTCGAGCAACAGCTAGCAGCTAGCAGCTAGCAGCTAGCAGCTAG 1896
DB 1426 GTAGATAGATCGGAGAGTCCCAACAGCTAGCAACAGCTAGCACTAGCACTAGCACTAG 1485

QY 1897 GGAACCTGAGAGTTCTCTGCTTACGCTTACGCTTACGACAGCTAGCAGCTAGCAGCTAG 1956
DB 1486 GGAACCTGAGAGTTCTCTGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACG 1545

QY 1957 AAGCGCTGCTCTACTACGG 1976
DB 1546 ACGGCTAAAGTTTATATGG 1565

RESULT 8
US-09-445-472-2
; Sequence 2, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: NORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyoko
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/05/445.472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
us-09-445-472-2

Query Match 42.3%; Score 837; DB 4; Length 1236;
Best Local Similarity 80.1%; Pred. No. 2.5e-175;
Matches 984; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

QY 1477 GACTACTCTACCGCTACTACGCTTCCGAGAGGCTCGCTACTACACCCGACGCC 1536
DB 1081 GACTACTCTTACCGCTACTACTATGATTCGAAGGTTGGTTATTACACCCACTGAT 1140
QY 1537 GGAACCTGGACGGTCAAGGTCGCTACAGGCGCGGCGGAGCTACAGGTGACGTC 1596
DB 1141 GGAACATGCAATTAAGGTTGTAAGCTACAGCGGAAGTCCAAACTTCAGTAGATGTG 1200
QY 1597 GTCAGCGAGCGGAGCTCAGCCAGTCCGG 1625
DB 1301 GTAAGTGATGGTTCCCTTTCCACAGCTGG 1329

RESULT 9
US-08-750-532-7
; Sequence 7, Application US/08750532
; Patent No. 5736339
; GENERAL INFORMATION:
; APPLICANT: NITTA, Masanori
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MORISHITA, Mio
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikudoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 Seventh Street N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,532
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JF95/01095
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1994/130236
; FILING DATE: 13-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1994/173912
; FILING DATE: 28-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: NITTA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 898 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-750-532-7

Query Match 33.9%; Score 569.8; DB 1; Length 898;
Best Local Similarity 86.3%; Pred. No. 1.4e-138;
Matches 780; Conservative 0; Mismatches 118; Indels 6; Gaps 4;

QY 523 GATCTCCAGGAAAGTAATTTGGTGGGTAGATTTTGTCAATGTTAGGAGTTATCCATC 582
DB 1021 TGGGACAAATGCAATAGGACCTTGATCTTTACCTACCATGATCCCAATGGAACACAGGTT 1080

QY 397 GCAGATTAGAGGACTGGATGAGTCTGACGCTCAAGTTATGCAACTTTAGCTTTGCGAC 456
DB 1 GCAGAAATTAGAAGGACTGGATGAGTCTGCGAGCTCAAGTTATGCAACTTTAGCTTTGGAAC 60
QY 457 TTGGGATATGATGGTTCTGGAATTCACATATAGGATTAATTCACACTGGATTGACGCTTCT 516
DB 61 TTGGGATATGATGGTTCTGGAATTCACATATAGGATTAATTCACACTGGATTGACGCTTCT 120
QY 517 CATCCAGATCTCCAGGAAAGTAATTTGGTGGGTAGATTTTCTCAATGGTAGGAGTTAT 576
DB 121 CATCCAGATCTCCAGGAAAGTAATTTGGTGGGTAGATTTTCTCAATGGTAGGAGTTAT 180
QY 577 CCATAGATGACATGACATGGAATCATGTAGCTTCAATAGCAGCTGGTAGTGGAGCA 636
DB 181 CCATAGATGACATGGAATCATGTAGCTTCAATAGCAGCTGGTAGTGGAGCA 240
QY 637 GCAAGTAATGGCAATGACAGGGAATGGCTCCAGGAGCTAAGCTGGCGGGAATTAAGTT 696
DB 241 GCAAGTAATGGCAATGACAGGGAATGGCTCCAGGAGCTAAGCTGGCGGGAATTAAGTT 300
QY 697 CTAGGTCCGATGGTTCTGGAAGCATATCTACTATAATTAAGGAGTGTAGTGGCGGCTT 756
DB 301 CTAGGTCCGATGGTTCTGGAAGCATATCTACTATAATTAAGGAGTGTAGTGGCGGCTT 360
QY 757 CATACAAAGTAAGTACGGAATTAAGTTCATTAATCTTCTTCTGTTCAAGCCAGAGC 816
DB 361 CATACAAAGTAAGTACGGAATTAAGTTCATTAATCTTCTTCTGTTCAAGCCAGAGC 420
QY 817 TCCGAGCAAGCACTCCCTCAGTCAGGCGCTCACACGCTGGGAGCCCGGTATAGTA 876
DB 421 TCAGATGCTACTGAGGCTCTAAGTCAGGCTGTAAATGAGCGGTGCGATGCTGGATAGTT 480
QY 877 GTCTGCTGCTCCGCGGCAACAGCGGCGCGAACACTACCGTCTGCTCACCAGCCGCC 936
DB 481 GTCTGCTGCTCCGCGGCAACAGCGTGGACCTAACAGTATACAAATTCGATGATAGTT 540
QY 937 GCGAGCAAGTCTATACGCTGGTGCAGTTGACAGCAAGCAACATCGCCAGCTTCTCC 996
DB 541 GCAGCAAGTCTATACGCTGGTGCAGTTGACAGCAAGCAACATCGCTGATGATTAACAAGCTCTCA 600
QY 997 AGCAGGAGCAAGCAAGCGGAGGAGCTCAAGCCGAGTGGTGGCGGCTTGGCGGCTCATCCTC 1056
DB 601 AGCAGGAGCAAGCAAGCGGAGGAGCTTAAAGCCTTAAGCCTGAGTTGTTCTCCAGCAAGCTG 660
QY 1057 ATCATAGCCCGCGCGCCAGCGAACCAGCATGGGCAACCCCGATTAAGAGCTACTACAGC 1116
DB 661 ATCATAGCCCGCGCGCCAGCGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 720
QY 1117 AAGGCTCTGGAACCAAGTGGCCACCGCGAGTTTGGGCGGTTGGCGGCTTGGCGGCTCATCCTC 1176
DB 721 AAGGCTCTGGAACCAAGTGGCCACCTCCACAGTCTGCTAGCTGTTGTTGCAAGCTCTTGTCTC 780
QY 1177 CAGGCGCCACCGAGCTGGACCCCGGACAAAGGTGAGACCGGCTCATCGAGCGCGGAC 1236
DB 781 CAGGCGCCACCGAGCTGGACCTCCAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 840
QY 1237 ATAGTTCGCGCGCGGAGGATAGCGGACATCCCTTACGGTTCGGGTAGGTTGACGCTCTAC 1296
DB 841 ATAGTTCGCGCGCGGAGGATAGCGGATAGCGGATAGCGGATAGCGGATAGCGGATAGCGGAT 900
QY 1297 AAGGCGCATAGTACGAGCAAGTACCGCAAGCTCACCTTCCAGCGGCTCGCTCGCGGCAAG 1356
DB 901 AAGGCGCATAGTACGAGCAAGTACCGCAAGCTCACCTTCCAGCGGCTCGCTCGCGGCAAG 960
QY 1357 GGAAGCGCCACCCACACCTTCGACGCTGACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1416
DB 961 GGAAGCGCCACCCACACCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
QY 1417 TGGGACACGGGCTGACGAGCATGACCTCTACCTCTACGACCCCGGCAAGCGAGGTT 1476
DB 1021 TGGGACACGGGCTGACGAGCATGACCTCTACCTCTACGACCCCAATGGAACACAGGTT 1080

Db 1 GATCTGAAGGGCAAGGTCATAGGCTGCTACGACGCCGTCACAGCGAGGTGCGACGCCCTAC 60
Qy 583 GATGACATGACATGGAAGTCATGCTGCTCAATACAGCTGCTGCTGAGGAGCAGAGT 642
Db 61 GATGACAGGAGGACGGAACCCAGTTCGGGTATCGTTCGCGAAGCCGAGCG---TT 117
Qy 643 AATGCAAGTACAGCAAGTGGCTCCAGAGCTAGCTGCGCGGAATTAAGTTTCTAGGT 702
Db 118 AACTCCAGTACATAGCGGTGCGCCCGCGCGGAGCTCGTGGGTCAAGGTTCTCGGT 177
Qy 703 GCGATGTTCTGGAAGCATATCTACTATAATTAAGGAGGTTGAGTGGGCGCTTGTATAAC 762
Db 178 GCGAGCGGTTCGGAGCGCTCCACCATCATCGCGGTGTGACTGGTGTCTCCAGAC 237
Qy 763 AAAGATAAGTACGAATTAAGTTCATTAATCTTCTTCTGTTGTTCAAGCAGAGCTCGGAC 822
Db 238 AAGGACAAGTACGGATAAGGTGTCATCACTCTCCCTCGGTCTCCAGAGCTCCGAC 297
Qy 823 GGAACCGACTCCCTCAGTCAAGCGCTCAACAGCGCTGGAGCGCGGTATAGTCTGC 882
Db 298 GGAACCGACTCCCTCAGTCAAGCGCTCAACAGCGCTGGAGCGCGGTATAGTCTGC 357
Qy 883 CTCGCGCGCGGACAGCGCGCGGCGGAAACCTACACCGTCCGCTCAACCGCGCGCGGAGC 942
Db 358 CTCGCGCGCGGACAGCGCGCGGCGGAAACCTACACCGTCCGCTCAACCGCGCGCGGAGC 417
Qy 943 AAGTCTATAACCTCGGTGCGTGTGAGTCAAGCAAGCAAGTCAAGTCAAGTCAAGTCAAG 1002
Db 418 AAGTCTATAACCTCGGTGCGTGTGAGTCAAGCAAGCAAGTCAAGTCAAGTCAAGTCAAG 477
Qy 1003 GACCGGACCGGACGAGGCTCAAGCGGAGTCTGCGCGCGCGGTGACATCATATA 1062
Db 478 GACCGGACCGGAGGCTCAAGCGGAGTCTGCGCGCGCGGTGACATCATATA 537
Qy 1063 GCGCGCGCGGACGAGGCAACCATGAGCGGCGCGGTAACGACTACTACACCAAGGCC 1122
Db 538 GCGCGCGCGGACGAGGCAACCATGAGCGGCGGTAACGACTACTACACCAAGGCC 597
Qy 1123 TCTGGAACGAGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1182
Db 598 TCTGGAACGAGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 657
Qy 1183 CACCGAGTCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1242
Db 658 CACCGAGTCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 716
Qy 1243 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1302
Db 717 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 776
Qy 1303 ATCAATAGAGGAGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1362
Db 777 ATCAATAGAGGAGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 836
Qy 1363 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1422
Db 837 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 894
Qy 1423 ACGG 1426
Db 895 ACGG 898

RESULT 10

US-08-894-8188-26
Sequence 26, Application US/088948188
Patent No. 6261822

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo

APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Kunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894, 8188
FILING DATE: 20-May-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-Nov-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-Dec-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25, 618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 507 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-8188-26

Query Match 25.6%; Score 507; DB 4; Length 507;
Best Local Similarity 100.0%; Pred. No. 7.7e-103;
Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 312 ACTAGTCATCTCAGGTTTAAACAGGGGTAAGCTTAAGCTTTCAGCTGTTAGGTTTATCCA 371
Db 1 ACTAGTCATCTCAGGTTTAAACAGGGGTAAGCTTAAGCTTTCAGCTGTTAGGTTTATCCA 60
Qy 372 GGAAGACTACAAAGTTACAGTTTACAGCAATTAAGAGGACTGGATGAGTCA 431
Db 61 GGAAGACTACAAAGTTACAGTTTACAGCAATTAAGAGGACTGGATGAGTCA 130
Qy 432 AGTTATGCACTTACGTTTGGAACTTGGATATGATGTTCTGGAATCACAATAGGAAT 491
Db 121 AGTTATGCACTTACGTTTGGAACTTGGATATGATGTTCTGGAATCACAATAGGAAT 180
Qy 492 AATTGACACTGGAATGACGCTTCTCATCCAGATCTCCAAAGGAAAGTAATTTGGTGGGT 551
Db 181 AATTGACACTGGAATGACGCTTCTCATCCAGATCTCCAAAGGAAAGTAATTTGGTGGGT 240
Qy 552 AGATTTTGTCAATGTTAGGATTTATCCATAGGATGAGTCAAGTCAATGAGTCAATGAGC 611
Db 241 AGATTTTGTCAATGTTAGGATTTATCCATAGGATGAGTCAAGTCAATGAGTCAATGAGC 300
Qy 612 TTCAATAGCAGCTGCTGCTGAGAGCAAGTAAATGGCAAGTACAAAGGAATGGTCCAGG 671
Db 301 TTCAATAGCAGCTGCTGCTGAGAGCAAGTAAATGGCAAGTACAAAGGAATGGTCCAGG 360
Qy 672 AGCTAAGCTGCGGGAAATTAAGGTTCTAGGTGCGGATGCTTCTGCAACCATATCTACTAT 731
Db 361 AGCTAAGCTGCGGGAAATTAAGGTTCTAGGTGCGGATGCTTCTGCAACCATATCTACTAT 420

4 CACGGAACCTACGTCGGGGAACAGTTGCCGGAACAGGACGCG---TTAACTCCCACTAC 60
 655 AAGGGAATGCTCCAGAGCTAAGCTGCGGGAATTAAGTTCTTAGGTGCGGATGCTTCT 714
 61 ATAGCGGTGCGCCCGCGCGGAGCTGCTGGTGTCAAGTTCTCGGTGCGGACGCTTGG 120
 715 GGAAGCATATCTACTATAATTAAGGAGTTGAGTGGCGGTGATTAACAAGATAAGTAC 774
 121 GGAAGCTCTCCACCATCATCGCGGTGTTGACTGGGTGCTCCAGAACAGGATAAGTAC 180
 775 GGAATTAAGGTCTATTATCTTCTCTGTTTCAAGCCAGAGCTCCGAGCAACCGACTCC 834
 181 GGAATTAAGGTCTATTATCTTCTCTGTTTCAAGCCAGAGCTCCGAGCAACCGACTCC 240
 835 CTCAGTCAAGCGCTCAACAGCGCTGGGAGCGGTATAGTCTGCTGCGCGCGCGGC 894
 241 CTCAGTCAAGCGCTCAACAGCGCTGGGAGCGGTATAGTCTGCTGCGCGCGGC 300
 895 AACAGCGCGCGGAAACCTACACCTGACCGCTGCGGTACCGCGCGCGCGCGCGGC 954
 301 AACAGCGCGCGGAAACCTACACCTGACCGCTGCGGTACCGCGCGCGCGCGGC 360
 955 GTCGGTGCACTGACAGCAACAGCAATCGCCAGCTTCTCCAGCAGGGAGCGCGGC 1014
 361 GTCGGTGCACTGACAGCAACAGCAATCGCCAGCTTCTCCAGCAGGGAGCGCGGC 420
 1015 GACGGAAGCTCAAGCGGAGTGTGCGCGCGCGGTGACATCATAGCCCGCGCGGC 1074
 421 GACGGAAGCTCAAGCGGAGTGTGCGCGCGCGGTGACATCATAGCCCGCGCGGC 480
 1075 AGCGGAACCGATGGCGCGCGGTGACATCATAGCCCGCGCGGTGACATCATAGCCCGCGGC 540
 481 AGCGGAACCGATGGCGCGCGGTGACATCATAGCCCGCGCGGTGACATCATAGCCCGCGGC 540
 1135 ATGCGCACCGCGCGCGGTGACATCATAGCCCGCGCGGTGACATCATAGCCCGCGGC 1157
 541 ATGCGCACCGCGCGCGGTGACATCATAGCCCGCGCGGTGACATCATAGCCCGCGGC 563

RESULT 12
 US-08-894-818B-14
 ; Sequence 14, Application US/08894818B
 ; Patent No. 6261822
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, HIKARU
 ; APPLICANT: MORISHITA, MIO
 ; APPLICANT: YAMAMOTO, KATSUHIKO
 ; APPLICANT: MITTA, MASAORI
 ; APPLICANT: ASADA, KIYOZO
 ; APPLICANT: TSUNASAWA, SUSUMU
 ; APPLICANT: KATO, IKUNOSHIN
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Browdy and Neimark
 ; STREET: 419 Seventh Street N.W., Ste. 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/894.818B
 ; FILING DATE: 20-MAY-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP95/03253
 ; FILING DATE: 07-NOV-1995
 ; PRIOR APPLICATION DATA:

732 AATTAAGGAGTTCAGTGGCGTGTGATAACAAGATAAGTACGGAATTAAGTCTATTAA 791
 421 AATTAAGGAGTTCAGTGGCGTGTGATAACAAGATAAGTACGGAATTAAGTCTATTAA 480
 792 TCTTCTCTCTGTTTCAAGCAGGCTC 818
 481 TCTTCTCTCTGTTTCAAGCAGGCTC 507

RESULT 11
 US-08-750-532-11
 ; Sequence 11, Application US/08750532
 ; Patent No. 5756339
 ; GENERAL INFORMATION:
 ; APPLICANT: MITTA, MASAORI
 ; APPLICANT: YAMAMOTO, KATSUHIKO
 ; APPLICANT: MORISHITA, MIO
 ; APPLICANT: ASADA, KIYOZO
 ; APPLICANT: TSUNASAWA, SUSUMU
 ; APPLICANT: KATO, IKUNOSHIN
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
 ; STREET: 419 Seventh Street N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/750,532
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP95/01095
 ; FILING DATE: 05-JUN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 1994/130236
 ; FILING DATE: 13-JUN-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 1994/173912
 ; FILING DATE: 26-JUL-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BROWDY, ROGER L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: MITTA-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 628-5197
 ; TELEFAX: (202) 737-3528
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 564 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..564
 ; US-08-750-532-11

Query Match 21.6%; Score 427; DB 1; Length 564;
 Best Local Similarity 86.1%; Pred. No. 3.2e-85;
 Matches 485; Conservative 0; Mismatches 75; Indels 3; Gaps 1;
 595 CATGGAACCTATGATGTTCAATAGCAGCTGCTGAGCAGCAAGTAATGGCAATAC 654

APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 564 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
US-08-894-818B-14

Query Match 21.6%; Score 427; DB 4; Length 564;
Best Local Similarity 86.1%; Pred. No. 3.2e-85;
Matches 485; Conservative 0; Mismatches 75; Indels 3; Gaps 1;

QY 595 CATGGACTCATGTAGCTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGSCAAGTAC 654
DB 4 CACGGACTCAGTGGCGGGAACAGTTGCCGGAACAGCAGCG---TTAATCTCCAGTAC 60
QY 655 AAGGGAATGGCTCCAGGAGCTAAGCTAGCTGGGGGAATTAAGGTTCTAGTGTCCGATGTTCT 714
DB 61 ATAGGCTGCGCCCGCGCGGCAAGCTCGTGTCAAGGTTCTCGGTGCCGACGGTTCG 120
QY 715 GGAAGCATATCTACTATATTAAGGAGTTGAGTGGCCGCTTGTACAAAGATTAAGTAC 774
DB 121 GGAAGCTGCTCCAGCATATCGCGGTGTTGACTGGGTGCTCCAGCAACAGGATAAGTAC 180
QY 775 GGAATTAAGTCAATATCTTCTTGTTCGAGGCTCCGAGCAGCTCCGAGCAGCTCC 834
DB 181 GGAATTAAGTCAATATCTTCTTGTTCGAGGCTCCGAGCAGCTCCGAGCAGCTCC 240
QY 835 CTCAGTACGGCGGCTCAACAGCGCTGGAGCGCGGTATAGTCTGCTGCGCGCGCGC 894
DB 241 CTCAGTACGGCGGCTCAACAGCGCTGGAGCGCGGTATAGTCTGCTGCGCGCGCGC 300
QY 895 AACAGCGGCGGCAACACTACACCGTCCGCTACCCCGCGCGCGCGCGCGCGCGCG 954
DB 301 AACAGCGGCGGCAACACTACACCGTCCGCTACCCCGCGCGCGCGCGCGCGCGCG 360
QY 955 GTCCGTGAGTTGACAGCAACAGCATCCAGCTTCTCCAGAGGGGAGCGACCGCGG 1014
DB 361 GTCCGTGAGTTGACAGCAACAGCATCCAGCTTCTCCAGAGGGGAGCGACCGCGG 420
QY 1015 GAGCGAGGCTCAAGCGGGAAGTCTGCGCCCGCGCGGTGACATCATAGCCCGCGCG 1074
DB 421 CAGCGAGGCTCAAGCGGGAAGTCTGCGCCCGCGCGGTGACATCATAGCCCGCGCG 480
QY 1075 AGCGAAGCAGCTAGGCGACCGGATAAGGCTACTACCAAGCGGCTGTGGAACCG 1134
DB 481 AGCGAAGCAGCTAGGCGACCGGATAAGGCTACTACCAAGCGGCTGTGGAACCG 540
QY 1135 ATGGCCCGCGCGAGCTTTCGGG 1157
DB 541 ATGGCCCGCGCGAGCTTTCGGG 563

RESULT 13
US-08-894-818B-30
Sequence 30, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo

APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikuroshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 419 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-30

Query Match 17.3%; Score 341.6; DB 4; Length 419;
Best Local Similarity 96.4%; Pred. No. 1.8e-66;
Matches 350; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 ATGAAGGGCTGAAGCTCTCATATTAGTATTAGTCTAGGTTTGGTAGTGGAGC 60
DB 56 ATGAAGGGCTGAAGCTCTCATATTAGTATTAGTCTAGGTTTGGTAGTGGAGC 115
QY 61 GTAGCGGAGCTCCAGAGAGAAAGTTGACCAAGTAAGAACTTTGAGAAAGTATGTT 120
DB 116 GTAGCGGAGCTCCAGAGAGAAAGTTGTTCAAGTAAGAAAGTTGAGAAAGTATGTT 175
QY 121 CTGCTAAGCGGAGCTGTCAGAAAAATTCAAAAATTTGAATCTTAACGAGAAATCAGC 180
DB 176 CTGCTAAGCGGAGCTGTCAGAAAAATTCAAAAATTTGAATCTTAACGAGAAATCAGC 235
QY 181 ACAGTAATTTGATTTGAAACCATAGGGGAAAAAGAAATTCAGTAAGAGTTCTTGAGTTA 240
DB 236 ACAGTAATTTGATTTGTAACCATAGGGGAAAAAGAAATTCAGTAAGAGTTCTTGAGTTA 295
QY 241 ATGGGTGCAAAAGTATAGTGTATGTTACCATATTATATACCCGCAATAGCTGCCGATTTAAG 300
DB 296 ATGGGTGCAAAAGTATAGTGTATGTTACCATATTATATACCCGCAATAGCTGCCGATTTAAG 355
QY 301 GTTAGAGACTTACTACTCATCTCAGGTTTAAACAGGGGGTAAAGCTTTCAGGTGTT 360
DB 356 GTTAGAGACTTACTACTCATCTCAGGTTTAAACAGGGGGTAAAGCTTTCAGGTGTT 415
QY 361 AGGT 364
DB 416 AGGT 419

us-09-841-553-6.rni

Thu Nov 7 10:12:17 2002

RESULT 14

US-08-894-818B-17
Sequence 17, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAKA, Susumu
APPLICANT: KATO, Ikumehiko
TITLE OF INVENTION: HYPERHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1464 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA

US-08-894-818B-17

Query Match 10.0%; Score 197; DB 4; Length 1464;
Best Local Similarity 59.0%; Pred. NO. 1.8e-34;
Matches 429; Conservative 0; Mismatches 265; Indels 33; Gaps 4;
QY 107 AGAAGAAGTCTGCTCTCAAGCCAGGACTGTTCAGAAAAATTCAGAAATTCAGTCTTA 166
DB 756 AGAAGAAGTCTGCTCTCAAGCCAGGACTGTTCAGAAAAATTCAGAAATTCAGTCTTA 815
QY 167 AGAAGAAGTCTGCTCTCAAGCCAGGACTGTTCAGAAAAATTCAGAAATTCAGTCTTA 226
DB 816 AGAAGAAGTCTGCTCTCAAGCCAGGACTGTTCAGAAAAATTCAGAAATTCAGTCTTA 875
QY 227 GAGTCTCTGAGTAAATGGGTGCAAAAGTTAGGTATGTGTACCATATATACCCGCAATAG 286
DB 876 AGTACTGAGCTCATCGGGCCAGGCTCAAGTACTCTTACAGAAATATCCCTCTGCTCG 935
QY 287 CTCCCGATCTTAAAGTCTAGAGCTTACGTATCTCATCTCAGCTTACAGTCTTACAG 334
DB 935 CGGTAAATTAAGCCAGGACCTTCTGCTGATCGGGGCATCATAGACACAGGCTTACT 995

QY 335 GGGTAAGCTAAGCTTTTCAGGCTTAGGTTTATCCAGGAAGACTACAAAGTTTACAGTTT 394
DB 996 TCGGTAAACACAGGGCTCGGGCATAAAGTTTCATACAGGAGGATTACAGGTTTACAGTTT- 1054
QY 395 CAGCAGAATTAGAGGACTGGATGAGTCTCCAGCTCAAGTTATGGCAACTTACGTTTGA 454
DB 1055 -----GACGACGCCACTTCCGTCCTCCAGATAGGGGCCGATACCGTCTGGA 1100
QY 455 ACT---TGGGATATGATGTTCTGGAATCAAAATAGGAATATAGTACACTGGAAATTTGAG 511
DB 1101 ACTCCCTCGGCTACGACGGAAGCGGTGGTGTCCCATCTCGTATACCGGTATAGACG 1160
QY 512 CTTCTCATCCAGATCTCCAAGGAAAGTAAATGGGTGGGTAGATTTTGTCAATGGTGA 571
DB 1161 CGAACCCCGGATCTGAAGGCAAGGTCATAGGCTGGTACGACTCCGTCACGGCAGGT 1220
QY 572 GTTATCCATAGGATGACCATGACATGGAACATCATGTAGCTTCAATAGCAGCTGGTACTG 631
DB 1221 CGACCCCTACGATGACGAGGACACGACCCACGTTCCGGGTATCGTTGCCGGAACCG 1280
QY 632 GAGCAGCAAGTAATGCAAGTACAGGGAATGGCTCCAGGAGCTTACGCTGCGGGAATTA 691
DB 1281 GGAGCG---TTAACTCCCACTACATAGGCGTCCGCCGCCGCGGAGGCTCTCGGCGTCA 1337
QY 692 AGTCTTAGTGGCGGATGTTCTTGGAAAGCATATCTACTATAATTAAGGAGGTTGAGTGGG 751
DB 1338 AGGTTCTGGTTCGAGCGGTTCGGGAAGCGTCTCCACCATCATCGCGGCTGTGACTGGA 1397
QY 752 CCGTTGATAACAAAGATAAGTACGGAAATAGGTCATTAACTTTCTTCTGTTTCAAGCC 811
DB 1398 ACCTCCAGAACTAGGCAAGTACGGGATAAGGTCATCAACCTCTCCCTCGGCTCCCTCC 1457
QY 812 AGAGCTC 818
DB 1458 AGAGCTC 1464

RESULT 15

US-09-000-016-3
Sequence 3, Application US/09000016
Patent No. 6143541
GENERAL INFORMATION:
APPLICANT: AKIRA ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
TITLE OF INVENTION: ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200

